НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра інформаційної безпеки

«На правах рукопису»

УДК 004.056

«До захисту допущено»
В.о. завідувача кафедри
_____ М.В.Грайворонський
“___”_____2018 р.

Магістерська дисертація

на здобуття ступеня магістра

зі спеціальності: 125 Кібербезпека

на тему: Фільтрація Твіттер-стрічки у режимі реального часу за допомогою машинного навчання

Виконав (-ла): студент (-ка) 2 курсу, групи ФБ-71мп
(шифр групи)

Зацепін Олексій Артемович
(прізвище, ім’я, по батькові)

Науковий керівник к.т.н., доц. Родіонов Андрій Миколайович __ _____
(посада, науковий ступінь, вчене звання, прізвище та ініціали) (підпис)

Консультант _____ _____ _____
(назва розділу) (науковий ступінь, вчене звання, , прізвище, ініціали) (підпис)

Рецензент к.т.н., доцент ХНУРЕ Сінельнікова О.І. _____
(посада, науковий ступінь, вчене звання, прізвище та ініціали) (підпис)

Засвідчую, що у цій магістерській дисертації немає запозичень з праць інших авторів без відповідних посилань.
Студент _____
(підпис)

Київ – 2018 року

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра інформаційної безпеки

Рівень вищої освіти – другий (магістерський) за освітньо-професійною програмою
Спеціальність (спеціалізація) – 125 Кібербезпека («Системи і технології кібербезпеки)

ЗАТВЕРДЖУЮ
В.о. завідувача кафедри
_____ М.В.Грайворонський
(підпис)

«___»_____2018 р.

ЗАВДАННЯ
на магістерську дисертацію студенту

Зацепіну Олексію Артемовичу

1. Тема дисертації: Фільтрація Твіттер-стрічки у режимі реального часу за допомогою машинного навчання
науковий керівник дисертації к.т.н., доц. Родіонов Андрій Миколайович,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від «15» листопада 2018 р. № 4171-с

2. Термін подання студентом дисертації 12.12.2018 р.

3. Об'єкт дослідження _____
_____

4. Вихідні дані _____
_____
_____
_____

5. Перелік завдань, які потрібно розробити _____
_____
_____
_____
_____
_____

6. Орієнтовний перелік ілюстративного матеріалу _____
_____
_____

7. Орієнтовний перелік публікацій _____
_____

8. Консультанти розділів дисертації*

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

9. Дата видачі завдання _____

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка

Студент _____ (підпис) _____ (ініціали, прізвище)

Науковий керівник дисертації _____ (підпис) _____ (ініціали, прізвище)

* Консультантом не може бути зазначено наукового керівника магістерської дисертації.

РЕФЕРАТ

Робота обсягом 107 сторінок містить 62 ілюстрацій, 22 таблиці, додаток та 17 літературних джерел.

Метою даного дослідження є побудова методу фільтрації твіттер-стрічки від контенту, згенерованого ботами за допомогою алгоритмів машинного навчання.

Об'єктом дослідження є процес фільтрації на контент, згенерований ботами.

Предметом дослідження є моделі, методи алгоритми для визначення контенту, згенерований ботами.

Використовувались такі методи дослідження як : підбір існуючої літератури по обраній темі, її опрацювання та визначення ключових аспектів. Аналіз технічної документації. Структуризація та систематизація усіх даних зібраних в результаті вивчення існуючих матеріалів по тематиці роботи. Аналіз ключових параметрів та побудова моделі для класифікації даних та реалізації програми-фільтра.

Результати роботи можуть бути використані для створення програмного забезпечення для його подальшого використання в системах відображення твітер-стрічки в режимі реального часу.

ТВИТТЕР, АНАЛІЗ ДАНИХ, СОЦІАЛЬНІ МЕРЕЖІ, БОТИ, ФІЛЬТРАЦІЯ, МАШИННЕ НАВЧАННЯ, АНАЛІЗ НАСТРОЇВ.

ABSTRACT

The work of 107 pages contains 62 illustrations, 22 tables, an appendix and 17 literary sources.

The purpose of this study is to construct a Twitter filtering method for content generated by bots using machine learning algorithms.

The object of the study is the process of filtering the content generated by the bots.

The subject of the study is models, methods of algorithms for determining content generated by bots.

The following research methods were used: the selection of existing literature on the chosen topic, its elaboration and the definition of key aspects. Analysis of technical documentation. Structuring and systematization of all data collected on the study of existing materials on the subject of work. Analysis of key parameters and model construction for data classification and implementation of the program filter.

The results of the work can be used to create software for its further use in systems display tweeter tape in real time.

TWITTER, ANALYSIS OF DATA, SOCIAL NETWORKS, BOTS, FILTRATION, MACHINE LEARNING, SENTIMENT ANALYSIS.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ОС – операційна система.

Лог – інформація про процес виконання певних дій.

ПЗ – програмне забезпечення.

Твіт - одне повідомлення в соціальній мережі Твіттер.

Машинне навчання - це підгалузь штучного інтелекту в галузі інформатики, яка часто застосовує статистичні прийоми для вирішення проблем класифікації і регресії.

Бот - спеціальна програма, що виконує автоматично і / або за заданим розкладом будь-які дії через інтерфейси, призначені для людей.

ВСТУП

Сучасні технології та розвиток техніки призвели до створення великої кількості різного роду автоматизованих інформаційних систем, таких як соціальні мережі. Неправомірне перекручення, фальсифікація, знищення або розголошення певної частини інформації, так само як і дезорганізація процесів її обробки і передачі в інформаційних системах завдають серйозної матеріальної та моральної шкоди багатьом користувачам, які беруть участь в процесах автоматизованого інформаційного взаємодії та навпаки можуть бути значним важелем у політичному та соціальному аспекті людства. Важливі інтереси цих суб'єктів, як правило, полягають в тому, щоб певна частина інформації, що стосується їх економічних, політичних та інших сторін діяльності, конфіденційна комерційна і персональна інформація, була б постійно легко доступна і в той же час надійно захищена від неправомірного її використання : небажаного розголошення, фальсифікації, спаму, блокування або знищення. На даний момент існує дуже багато спаму, реклами та штучно створеного суб'єктивного контенту, що може впливати на суспільство.

Актуальність роботи є високою, через те, що під впливом розвитку соціальних мереж та інтернету загалом робить їх дуже впливовими на людство. Не дивлячись на зусилля Facebook, Twitter, Instagram на ринок все одно потрапляють потенційно небезпечні застосунки, які спотворюють публічні данні.

Метою данної роботи є побудова методу фільтрації твіттер-стрічки від контенту, згенерованого ботами за допомогою алгоритмів машинного навчання.

Об’єктом дослідження є процес фільтрації на контент, згенерований ботами.

Предметом дослідження є моделі, методи алгоритми для визначення контенту, згенерований ботами.

В результаті аналізу мети роботи були визначені наступні **завдання дослідження**:

- огляд та аналіз існуючих соціальних мереж, зокрема Твіттеру;
- аналіз поведінки програм-ботів;
- вибір методів машинного навчання для класифікації ботів;
- побудова моделі користувача та застосування методів машинного навчання для фільтрації штучного контенту;
- реалізація програмного рішення на основі обраної моделі та аналіз результатів його роботи.

Наукова новизна роботи полягає в побудові моделі фільтрації Твіттер-стрічки від штучного контенту на основі алгоритмів машинного навчання та аналізу тексту повідомлення.

Практичне значенням результатів є програмне застосування, що в режимі реального часу дозволяє фільтрувати Твіттер-стрічку від штучного контенту на основі машинного навчання та аналізу природних мов. Пропоноване рішення може використовуватись при побудові плагіну для браузеру.

1 АНАЛІЗ ПОВЕДІНКИ БОТІВ В СОЦІАЛЬНИХ МЕРЕЖАХ

1.1 Соціальні мережі

Соціальна мережа - це соціальна структура, яка складається з безлічі соціальних суб'єктів (таких як особи або організації), набори діадних зв'язків та інші соціальні взаємодії між суб'єктами. Перспектива соціальної мережі передбачає набір методів аналізу структури цілих соціальних сутностей, а також різноманітних теорій, що пояснюють шаблони, що

спостерігаються в цих структурах . Дослідження цих структур використовує аналіз соціальної мережі для визначення локальних та глобальних моделей, визначення впливових об'єктів та вивчення динаміки мережі. Соціальні медіа та соціальні мережі зосереджуються на двосторонній взаємодії, між сайтом (або людиною, яка працює на сайті) та людьми, які її читають чи використовують [1].

Соціальні мережі та їх аналіз є по суті міждисциплінарною академічною сферою, яка виникла з соціальної психології, соціології, статистики та теорії графів. Соціальна мережа є теоретичною побудовою, корисною в соціальних науках для вивчення відносин між окремими особами, групами, організаціями або навіть цілими товариствами. Цей термін використовується для опису соціальної структури, визначеної такими взаємодіями. Зв'язки, через які з'єднує будь-який певний соціальний блок, є зближення різних соціальних контактів цієї одиниці. Цей теоретичний підхід, обов'язково, є реляційним. Аксіома соціального підходу до розуміння соціальної взаємодії полягає в тому, що соціальні явища повинні в першу чергу задуматися і досліджуватися через властивості відносин між і всередині одиниць, а не властивостями самих цих одиниць. Таким чином, однією загальною критикою теорії соціальних мереж є те, що індивідуальне агентство часто ігнорується, хоча це може бути не на практиці. Саме тому, що ці мережеві конфігурації утворюють безліч різних типів взаємозв'язків, окремо або в комбінації, мережева аналітика корисна широкому колу дослідницьких підприємств. У соціальній науці ці сфери навчання включають в себе, але не обмежуючись ними, антропологію, біологію, комунікативні дослідження, економіку, географію, інформаційну науку, організаційні дослідження, соціальну психологію, соціологію та соціолінгвістику.

Комп'ютерні мережі у поєднанні із програмним забезпеченням соціальної мережі створюють нове середовище для соціальної взаємодії.

Відносини до комп'ютеризованої служби соціальних мереж можна характеризувати контекстом, напрямком та силою. Зміст відношення відноситься до ресурсу, який обмінюється. У комунікаційному контексті, пов'язаному з комп'ютером, соціальні пари обмінюються різними видами інформації, включаючи передачу файлів даних або комп'ютерної програми, а також надання емоційної підтримки або організації зустрічі. Зі зростанням електронної торгівлі обмін інформацією також може відповідати обміном грошима, товарами або послугами у "реальному" світі. Методи аналізу соціальних мереж стали важливими для вивчення цих типів комп'ютерної комунікації. Крім того, величезні розміри та нестабільність соціальних мереж призвели до появи нових мережевих показників. Основна проблема, пов'язана з мережами, витягнутою з соціальних мереж, полягає у відсутності надійності мережевих показників із відсутністю даних [2].

1.1.1 Вплив соціальних мереж на людей

У дослідженні, проведеному дослідженнями Інтернет-проекту Pew, було показано, що 67% дорослих онлайн-користувачів використовують соціальні мережі. Ці дані описують той факт, що соціальні мережі можуть бути використані для поліпшення освіти дорослих і студентів, оскільки вони вже мають певний вплив на учнів. Основна перевага полягає в тому, що студенти можуть додатково вивчати теми, які їм цікавлять, використовуючи онлайнові соціальні мережі, оскільки вони можуть мати обмежені ресурси та час у школі. Окрім того, для дорослих з унікальними інтересами нелегко знайти друзів з тим самим інтересом спілкуватися та обговорювати. Однак соціальні мережі пов'язують цих людей з такими ж пристрастями, де вони можуть взаємодіяти з людьми з різних частин світу. З іншого боку, вчителі можуть скористатися перевагами можливостей соціальних мереж учнів для створення дискусійних форумів, класних

блогів та онлайн-навчання. Така співпраця між студентами та викладачами, що діють соціальними мережами як середовище, може допомогти учням отримати можливості для покриття знань у більш широкій області та зацікавлення. Школи також використовують соціальні мережі як проміжну, щоб підтримувати зв'язок зі студентами. Деякі мережі, такі як Facebook, Moodle, Secondlife, Digg та інші мережі, часто використовуються вчителями для спілкування зі студентами та проведення дискусій поза аудиторією. Отже, цілком очевидно, що для впливу соціальних мереж на освіту спостерігаються кілька переваг [3].

Соціальні мережі відбуваються онлайн, де люди зустрічаються та діляться ідеями, рекомендаціями та досвідом. Таким чином, це схоже на безперервне спілкування з широкою аудиторією за допомогою різних платформ або сайтів. Це один з факторів, яким організації цікавлять соціальні медіа, оскільки вони можуть отримувати різні пропозиції та зворотні зв'язки від людей. Багато компаній використовують соціальні мережі, такі як Twitter і Facebook, для спілкування з клієнтами та потенційними клієнтами. Це призводить до золотої можливості для шукачів роботи, щоб дізнатися більше про організацію та легко спілкуватися з людьми, які там працюють. Шукачі роботодавців можуть ознайомитися з тими, хто працює через ці соціальні медіа, і як тільки вони досягають сильної присутності в цих мережах, стає реальним домогтися людей з правом наймати працівника. Різні види соціальних мереж передають ефективні способи пошуку роботи. Ви можете знайти інформацію про компанію в Google, просто набравши назву цієї компанії. Facebook, Twitter і FourSquare - відмінні соціальні мережі, які дозволять взаємодіяти з людьми, які працюють в організації. Якщо пощастить, то можна зустрітись з агентом-рекрутером і провести розмову та обговорення, які можуть збільшити шанс отримати роботу. Крім того, профіль можна розподілити в соціальних мережах, що збільшує

ймовірність отримання роботи. Маріам Салпітер, засновник кар'єри Кеппі, стверджує, що "Створення онлайн-присутності дозволяє наймати менеджера, рекрутера, колеги та друзів, щоб більше дізнатися про вас, про те, що ви пропонуєте та про те, що ви хочете. Це спосіб залучення робочих місць замість того, щоб витрачати час на пошуки робочих місць. "Однією з найважливіших моментів, щоб вказати вплив соціальних мереж на бізнес-напрям, є маркетинг соціальних мереж. Діючі соціальними мережами як крок, маркетинг соціальних мереж може отримати багато переваг, включаючи створення відносин, створення брендів, рекламу, просування по службі тощо. Таким чином, можна зробити висновок, що маркетинг соціальних мереж пропонує декілька можливостей для підприємців, малого бізнесу, середніх компаній та великих корпорацій для побудови своїх брендів та бізнесу. Підкреслюючи переваги соціальних мереж, це не обов'язково означає, що в них немає ніяких недоліків. Нещодавно онлайн злочинність, також відома як кіберзлочинність, пропонує збільшити загрозу для всіх користувачів Інтернету. Це включає в себе сексуальну експлуатацію та кібер-залякування в Інтернеті. Однією з головних проблем, спрямованих на викорінення кіберзлочинів, є те, що важко визначити правопорушника, і практично неможливо вести постійний нагляд у такій широкій мережі. Однією з найбільш агресивних форм кіберзлочинності є сексуальна експлуатація в Інтернеті. Це включає в себе обмін порнографією, переконуючи секс і секс чат. У Сполучених Штатах Америки зареєстровано понад 665000 зареєстрованих осіб, які засуджені за статевими злочинами, згідно з дослідженням, за замовленням Національного центру зникнення та експлуатації дітей. Це означає, що один з кожних семи дітей підійшов до сексуального хижака в Інтернеті. Тобто 13% дітей користуються Інтернетом. Крім того, Центр з управління сексуальними преступниками (CSOM) зазначив, що середній сексуальний правопорушник ображається протягом 16 років, перш ніж його нарешті

потраплять. У цьому життєвий цикл він здійснив і в середньому 318 злочинів і порушив 110 жертв. Що стосується цих даних, то чітко видно, що сексуальна експлуатація в Інтернеті руйнує життя дітей, які використовують соціальні мережі.

Кібер-залякування відрізняється від залякування особистого лиця, оскільки неможливо легко виявити хуліганів, і вони мають почуття безпеки, що переконує їх, що вони не попадуть. Не знаючи тієї шкоди, яку вони заподіювали жертві, хулігани не відчувають ні провини, ні співпереживання. Кібернетичне вчинення також є формою кіберзлочинності, яка включає різні гілки. Найпоширеніший тип називається пригнобленням, де зловмисні та образливі повідомлення неодноразово надсилаються жертві. Інші види кишенькового хулігансу, такі як полум'я, маніфестація, висування себе за іншу особу, виїзд, хитрість та виключення також знаходяться в світі соціальних мереж. Найбільш важливим способом називається кібернетик, що створює переслідування зі значними загрозами і створює страх. Соціальні мережі, що використовуються як форма асистента в галузі освіти, також демонструють негативні наслідки для студентів. Одним з таких впливів є пристрасть до мереж. Опитування, проведене Міжнародним центром медіа та зв'язків з громадськістю (МКМПА) в університеті Меріленда, показало, що у віці до 25 років більше шансів на залежність від соціальних мереж, а дві третини студентів, які використовують соціальні мережі, вже проявляють певну залежність . П'ятдесят відсотків людей у віці від 25 до 35 років визнали, що вони настільки прив'язані до соціальних мереж, що вони навіть використовують їх у робочі години. Крім того, діти можуть також стати прихильниками соціальних мереж, якщо батьківського керівництва немає. Окрім негативних аспектів соціальних мереж, є деякі випадки, коли вони можуть позитивно впливати на життя людей у майбутньому. Одне з таких впливів полягає в тому, що люди отримують

більш зручні способи життя. Проведення зустрічей та дискусійних форумів за допомогою соціальних мереж дасть можливість зайнятим бізнесменам більше часу проводити зі своєю сім'єю. Інтернет-магазини скоротять час домогосподарств, коли вони можуть робити інші домашні господарські роботи. Крім того, студенти можуть мати онлайн-дні навчання, де вони можуть спілкуватися з людьми з різних куточків світу та поділитися своїми ідеями та дебатами. Наступним фактом є те, що соціальні мережі допомагають людям підтримувати зв'язок із розвиваючими країнами. Це включає в себе обмін останніми новинами, фондові ціни акцій та ціни на золото. Оскільки соціальні мережі стають доступними на мобільних телефонах в ці дні, люди завжди будуть проінформовані про поточні новини світу. Нарешті, але не менш важливо, деякі фахівці соціальних мереж хочуть перейти на етап спілкування, де він може замінити телефон. Поки що соціальні мережі, такі як Google, Facebook і Skype, стали основними засобами масової інформації для спілкування за кордоном. Вчені соціальних мереж вважають, що вони можуть створювати нові комунікаційні технології, які можуть замінити мобільні телефони в майбутньому. Хоча соціальні мережі можуть вести багато позитивних впливів, вони одночасно накладають негативні. Основний недолік полягає в тому, що люди можуть почати втрачати фізичне спілкування та взаємодію в реальному світі. Наприклад, два людини стають кращими друзями в Інтернеті, але вони не спілкуються один з одним у реальному світі. Крім того, не буде розмови між партнерами на роботі, оскільки системи контролюють комп'ютер, і вони можуть спілкуватися, просто набравши кілька слів у вікні чату. Другий вплив - це справа студентів та дітей. Хоча студенти можуть отримати знання, коли вони використовують соціальні мережі, вони також можуть зіткнутися з загрозами та небезпеками в Інтернеті. Більшість людей вважають, що кіберзлочинність буде збільшуватися в майбутньому,

оскільки злочинці можуть маніпулювати різними способами та методами злочинів у такому світі з більш ніж 300 сайтами соціальних мереж. З іншого боку, існує ймовірність того, що люди, особливо віком від 15 до 25 років, будуть більше залежні від соціальних мереж у майбутньому. Крім того, оскільки соціальні мережі починають зосереджувати свої веб-сайти на розважальних та комерційних рекламних роликах, а не на освіті, люди можуть витрачати більше часу на мережу замість того, щоб читати або виконувати фізичні вправи. Витрати в часі в соціальних мережах не є сприятливими для здоров'я, тому що, коли він використовує мережі, він чи вона нічого не робить, крім сидячи перед экраном або лежачи на дивані.

1.1.2 Аналітика використання соціальних мереж

Поширення соціальних мереж в усьому світі постійно зростає. У 2017 році 71 відсоток користувачів Інтернету були користувачами соціальних мереж, і очікується, що ці цифри зростатимуть. Соціальні мережі - це одна з найпопулярніших онлайн-заходів із високим рівнем зацікавленості користувачів та розширенням мобільних можливостей. Північна Америка посідає перше місце серед регіонів, де соціальні медіа дуже популярні, причому рівень проникнення соціальних мереж становить 66 відсотків. У 2016 році більше 81 відсотків населення Сполучених Штатів мали профіль соціальних мереж. На другому кварталі 2016 року користувачі США проводять понад 215 щотижневих хвилин у соціальних мережах через смартфон, 61 тиждень на ПК та 47 хвилин на тиждень в соціальних мережах через планшетні пристрої [4].

Зростання в усьому світі використання смартфонів і мобільних пристроїв відкрило можливості мобільних соціальних мереж з покращеними функціями, такими як служби на основі місцезнаходження, як Foursquare або Google Now. Більшість соціальних мереж також доступні як мобільні соціальні додатки, тоді як деякі мережі були оптимізовані для

перегляду через мобільний Інтернет, що дозволяє користувачам зручно отримувати доступ до візуальних веб-сайтів, таких як Tumblr або Pinterest, через планшет. З більш ніж 1,86 мільярдами активних користувачів щомісяця, соціальна мережа Facebook в даний час є лідером ринку з точки зору охоплення та масштабу. Сайт формує ландшафт соціальних мереж з моменту його запуску, і є важливим фактором у дискусіях про конфіденційність користувачів та диференціювання приватного та громадського онлайнового Інтернету.

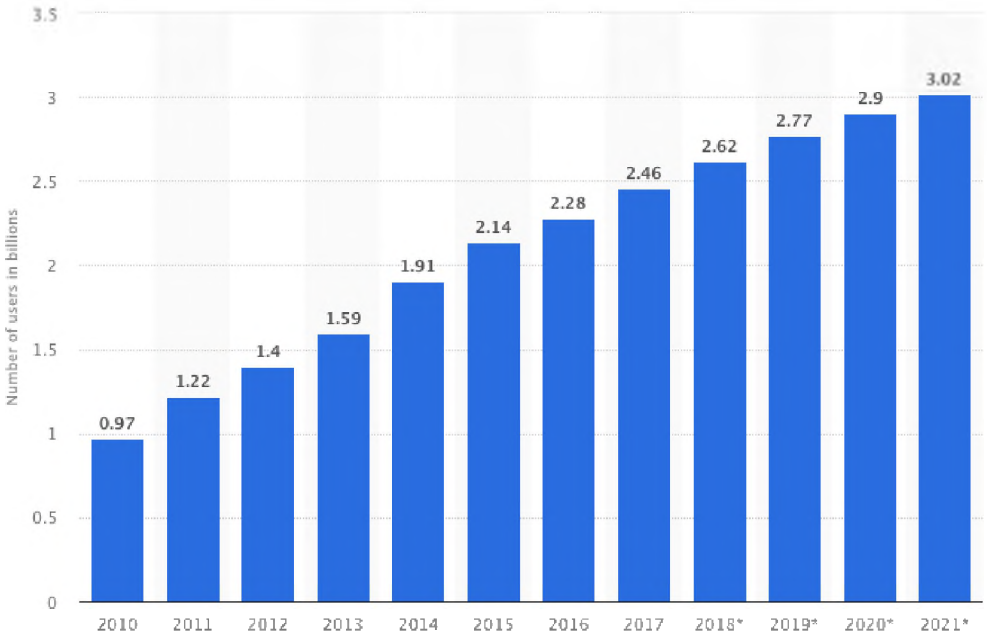

Рисунок 1.1 - Кількість користувачів в соціальних мережах

На рисунку 1 зображена кількість користувачів в соціальних мережах по роках [5]. Соціальні мережі не лише дозволяють користувачам спілкуватися за межами місцевих або соціальних кордонів, але також пропонують можливості для обміну користувальницьким вмістом, таким як фотографії та відео, а також функції, такі як соціальні ігри. Соціальна реклама та соціальні ігри є двома основними точками доходу для соціальних мереж.

Провідні соціальні мережі, як правило, доступні на декількох мовах і дозволяють користувачам спілкуватися з друзями або людьми з

географічними, політичними чи економічними межами. Близько 2 мільярдів користувачів Інтернету використовують соціальні мережі, і, як очікується, ці цифри зростатимуть як використання мобільного пристрою, а мобільні соціальні мережі все більше посилюються. Найпопулярніші соціальні мережі зазвичай показують велику кількість облікових записів користувачів або сильне зацікавленість користувачів. Наприклад, лідер на ринку Facebook став першою соціальною мережею, яка перевищила 1 мільярд активних користувачів щомісяця, тоді як нещодавно нова компанія Pinterest була найшвидшим самостійним сайтом для досягнення 10 мільйонів унікальних щомісячних відвідувачів. Більшість соціальних мереж із понад 100 мільйонами користувачів походять з США, однак європейські послуги, такі як VK або китайські соціальні мережі, Qzone та Renren також отримали популярність у своїх регіонах за рахунок локального контексту та змісту.

Споживачі соціальної мережі споживачами дуже різноманітні: такі платформи, як Facebook або Google+, цілком зосереджені на обміні друзями та сім'єю та постійно ведуть взаємодію через такі функції, як обмін фотографіями чи статусом, а також соціальні ігри. Інші соціальні мережі, такі як Tumblr або Twitter, стосуються швидкого спілкування та влучно називаються мікроблогами. Деякі соціальні мережі зосереджені на громаді; інші підкреслюють та відображають вміст, створений користувачем. Завдяки постійній присутності в житті своїх користувачів, соціальні мережі мають рішуче сильний соціальний вплив. Розмивання між автономним та віртуальним життям, а також концепція цифрової ідентичності та соціальних взаємодій у мережі - це деякі аспекти, які виникли в ході останніх обговорень.

У цій сфері домінують Facebook і YouTube, оскільки значна більшість дорослих США використовує кожен з цих сайтів. У той же час, молодші американці (особливо у віці від 18 до 24 років) виділяють широке

коло різноманітних платформ та часто використовують їх. Близько 78% людей віком від 18 до 24 років використовують Snapchat, а значна більшість цих користувачів (71%) відвідують платформу кілька разів на день. Подібним чином, 71% американців цієї вікової групи зараз використовують Instagram, а близько половини (45%) - користувачі Twitter.

Facebook залишається найпоширенішою платформою соціальних мереж за відносно здоровою ціною: приблизно 68% дорослих США зараз є користувачами Facebook. За винятком платформи для обміну відео YouTube, жоден з інших сайтів чи програм, які вимірюються в цьому опитуванні, не використовується більш ніж 40% американців [6]. Центр поцікавився використанням п'яти з цих платформ (Facebook, Twitter, Instagram, LinkedIn та Pinterest) у кількох попередніх опитуваннях про використання технологій. І в основному частка американців, які користуються кожним із цих послуг, подібна до того, що Центр знайшов у своєму попередньому опитуванні про використання соціальних мереж, який був проведений у квітні 2016 року. Найбільш відомим винятком є Instagram: 35% дорослих США зараз говорять вони використовують цю платформу, збільшившись на 7 процентних пунктів з 28%, які заявили, що вони зробили це в 2016 році. Також існують істотні відмінності у використанні соціальних мереж за віком. Близько 88% людей віком від 18 до 29 років вказують на те, що вони використовують будь-яку форму соціальних мереж. Ця частка становить 78% у віці від 30 до 49 років, до 64% у віці від 50 до 64 років і до 37% серед американців 65 років і старше.

У той же час існують помітні відмінності у використанні різних платформ соціальних мереж у молодому дорослому віці. Американці віком від 18 до 24 років набагато частіше використовують такі платформи, як Snapchat, Instagram і Twitter, навіть у порівнянні з тими, що знаходяться в середині до кінця 20-х років. Ці відмінності особливо помітні, коли мова

йде про Snapchat: 78% 18-24-річних людей є користувачами Snapchat, але серед 25-29 років ця частка падає до 54%.

За винятком тих, кого 65 років і старше, Facebook використовується більшістю американців у широкому діапазоні демографічних груп. Але інші платформи більш рішуче звертаються до певних підгруп населення. На додаток до вікових відмінностей у використанні таких сайтів, як Instagram та Snapchat, зазначені вище.

Середній американець використовує три з цих восьми соціальних платформ. Як і раніше, у попередніх опитуваннях щодо використання соціальних мереж існує значне співпадіння користувачів користувачів на різних сайтах, виміряних у цьому опитуванні. Зокрема, значна більшість користувачів кожної з цих соціальних платформ також вказують на те, що вони використовують Facebook і YouTube. Але ця "взаємність" поширюється і на інші сайти. Наприклад, приблизно три чверті користувачів Twitter (73%) та Snapchat (77%) також вказують на те, що вони використовують Instagram.

	Use Twitter	Use Instagram	Use Facebook	Use Snapchat	Use YouTube	Use WhatsApp	Use Pinterest	Use LinkedIn
Twitter	–	73%	90%	54%	95%	35%	49%	50%
Instagram	50	–	91	60	95	35	47	41
Facebook	32	47	–	35	87	27	37	33
Snapchat	48	77	89	–	95	33	44	37
YouTube	31	45	81	35	–	28	36	32
WhatsApp	38	55	85	40	92	–	33	40
Pinterest	41	56	89	41	92	25	–	42
LinkedIn	47	57	90	40	94	35	49	–

Source: Survey conducted Jan. 3-10, 2018.
"Social Media Use in 2018"
PEW RESEARCH CENTER

90% of LinkedIn users also use Facebook

Рисунок 1.2 - Кількість однакових користувачів у різних мережах.

Це співпадіння, що зображені на рисунку 2 в цілому свідчить про те, що багато американців використовують кілька соціальних платформ. Приблизно три чверті громадськості (73%) використовує більше ніж одну з восьми платформ [7], виміряних у цьому опитуванні, а типовий

(середній) американець використовує три з цих сайтів. Як і слід було очікувати, молодші дорослі користувачі, як правило, використовують більший спектр соціальних медіа-платформ. Середній вік від 18 до 29 років використовує чотири з цих платформ, але цей показник зменшується до трьох серед 30- 49-річних, до двох серед 50-64-річних та до одного з 65-ти і старше

1.1.3 Твіттер

Twitter - це служба соціальних мереж та мікроблогів, яка дозволяє зареєстрованим користувачам читати та публікувати короткі повідомлення, так звані твіти, була запущена в березні 2006 року. Повідомлення Twitter обмежуються 280 символами, а користувачі також можуть завантажувати фотографії або короткі відеоролики. Твіти надсилаються в загальнодоступний профіль або можуть надсилатися як прямі повідомлення іншим користувачам. Twitter - одна з найпопулярніших соціальних мереж у всьому світі. Частина звернення - це здатність користувачів стежити за будь-яким іншим користувачем із загальнодоступним профілем, що дозволяє користувачам взаємодіяти зі знаменитостями, які регулярно розміщують на сайті соціальної мережі. На сьогоднішній день найбільш популярною людиною в Twitter є співачка Кеті Перрі з понад 108 мільйонами послідовників.

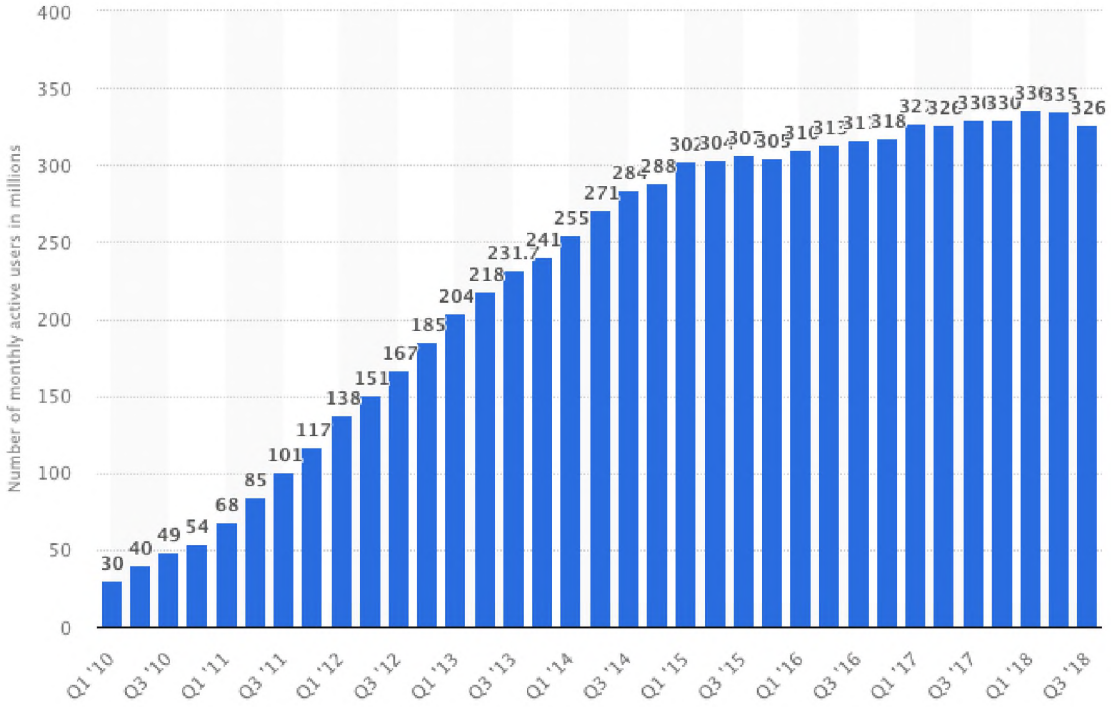

Рисунок 1.3 - Кількість користувачів Твіттеру.

Twitter також став важливим каналом комунікацій урядам та главам держав - колишній президент США Барак Обама заявив, що це перше місце в поглядах на послідовників Twitter, з прем'єр-міністром Індії Нарандрою Моді та президентом Туреччини Реджепом Тайіпом Ердоганом, який займає друге і третє місце відповідно. У світлі зростання послуг у сфері електронного уряду в усьому світі це не дивно. Незважаючи на стійкий ріст доходів, прибуток компанії у 2017 році склав 2,44 млрд. Доларів США, зменшився з 2,5 млрд. У попередньому фінансовому році, - Twitter ще не звітував про позитивний річний чистий дохід. У 2017 році щорічний чистий збиток склав 108 млн. Доларів США. Статистика на рисунку №3 показує часову шкалу з кількістю щомісячних активних користувачів Twitter у всьому світі. Станом на третій квартал 2018 р. Мікроблоги обслуговувалися в середньому 326 мільйонів щомісячно активних користувачів.

Також твіттер є платформою де користувачі найбільше читають новини.

	2013	2016	2017
Twitter	52%	59%	74%
reddit	62%	70%	68%
Facebook	47%	66%	68%
Tumblr	29%	31%	39%
YouTube	20%	21%	32%
Snapchat	-	17%	29%
Instagram	13%	23%	27%
LinkedIn	13%	19%	23%
WhatsApp	-	-	23%

Note: ages 18+; among users of each platform
Source: Pew Research Center, "News Use Across Social Media Platforms 2017," Sep 9, 2017

Рисунок 1.4 - Відсоток пізнання новин з соціальних мереж.

Структура повідомлення в соціальній мережі Твіттер:

Таблиця 1.1 - Структура повідомлення.

Поле	Означення
created_at	Час UTC, коли цей твіт був створений.
id	Ціле представлення унікального ідентифікатора для цього твіту.
id_str	Символьне представлення унікального ідентифікатора для цього твіту.
text	Фактичний UTF-8 текст оновлення статусу.
user	Користувач, який опублікував цей твіт.
entities	Об'єкти, які були проаналізовані з тексту твіту. (Хештеги, медіа)

Кінець таблиці 1.1

place	Коли вони присутні, вказує на те, що твіт пов'язано (але не обов'язково) з місцем

quote_count	Показує приблизно скільки разів цей твіттер цитував користувачів Twitter.
reply_count	Скільки разів на цей твір відповіли.
retweet_count	Кількість разів, коли цей твіт був ретвітнутий.
favorite_count	Показує приблизно скільки разів цей твіттер сподобався користувачам Twitter.

У таблиці 1 було відображено структуру повідомлення (твіта) у соціальній мережі твіттер.

1.2 Боти

Бот - це спеціальна програма, що виконує автоматично і / або за заданим розкладом будь-які дії через інтерфейси, призначені для людей [8].

1.2.1 Види ботів

- Веб-сканер

Веб-сканер, який іноді називають павуком або spiderbot і часто скорочується до сканера, - це інтернет-бот, який систематично переглядає Всесвітню мережу, як правило, для індексації веб-сайтів (spidering). Веб-пошукові системи та деякі інші сайти використовують веб-сканування або програмне забезпечення для спамерів, щоб оновити свій веб-вміст або індекси веб-контенту інших сайтів. Веб-сканери копіюють сторінки для обробки пошуковою системою, яка індексує завантажені сторінки, щоб користувачі могли здійснювати пошук більш ефективно. Скануючі користувачі витрачають ресурси на відвідані системи та часто відвідують сайти без схвалення. Проблеми графіка, завантаження та "ввічливості"

вступають у гру, коли досягаються великі збірки сторінок. Механізми існують для загальнодоступних сайтів, які не хочуть сканувати, щоб зробити це відомим сканеру. Наприклад, у тому числі файл robots.txt може вимагати від ботів індексувати лише частини веб-сайту або взагалі нічого. Кількість веб-сторінок дуже велика; навіть найбільші сканери не мають повного індексу. З цієї причини пошукові машини намагалися дати відповідні результати пошуку в перші роки Всесвітньої павутини до 2000 року. Сьогодні відповідні результати даються практично миттєво. Скануючі користувачі можуть перевіряти гіперпосилання та код HTML. Вони також можуть бути використані для веб-скребків (див. Також керування даними програмування).

- Інтернет бот

Інтернет-бот, також відомий як веб-робот, WWW робот або просто бот, являє собою програмне додаток, що управляє автоматизованими завданнями (скриптами) через Інтернет. Як правило, боти виконують завдання, які є простими та структурно повторюваними, при значно більш високій швидкості, ніж можна було б тільки для людини. Найбільше використання ботів - це веб-спрейер (веб-сканер), в якому автоматичний скрипт отримує, аналізує та надсилає інформацію з веб-серверів у багато разів швидкість роботи людини. Більше половини всього веб-трафіку складається з ботів. Зусилля серверів, які розміщують веб-сайти для протидії ботам, різняться. Сервери можуть обирати правила для поведінки інтернет-ботів, застосувавши файл robots.txt: цей файл є просто текстом, що визначає правила поведінки бота на цьому сервері. Будь-який бот, який не дотримується цих правил під час взаємодії з (або "spidering") будь-яким сервером, теоретично, повинен бути позбавлений доступу до веб-сайту, на який це було порушено, або його видалити. Якщо єдиним правилом, реалізованим на сервері, є текстовий файл, не пов'язаний із програмою / програмним забезпеченням / додатком, то дотримання цих правил є цілком

добровільним - насправді неможливо застосувати ці правила або навіть забезпечити, щоб творець бота або виконавець визнає або навіть читає вміст файлу robots.txt. Деякі боти "добре" - наприклад, павуки пошукової системи, тоді як інші можуть використовуватися для запуску шкідливих і суворих нападів, зокрема, у політичних кампаніях.

- Чатбот

Чатбот (також відомий як smartbots, talkbot, chatterbot, Bot, IM бот, інтерактивний агент, діалоговий інтерфейс або штучний розмовний об'єкт) - це комп'ютерна програма або штучний інтелект, який веде розмову через слухові чи текстові методи. Такі програми часто розроблені, щоб переконливо імітувати, як людина буде вести себе як розмовний партнер, тим самим пройшов тест Тьюринга. Чабаботи зазвичай використовуються в месенджерах для різних практичних цілей, включаючи онлайн допомогу клієнтам або купівлю-продаж. Деякі чатботи використовують складну систему обробки природних мов, але багато з них просто сканують для ключових слів у вхідному документі, після чого надсилайте відповідь з найбільш відповідних ключових слів або найбільш схожих шаблонів формулювання з бази даних. Термін "Chatter Bot" спочатку було створено Майкла Маулдіна у 1994 році, щоб описати ці розмовні програми. Сьогодні більшість обговорень доступні через віртуальних помічників, таких як Google Assistant та Amazon Alexa, за допомогою програм обміну повідомленнями, таких як Facebook Messenger або WeChat, або через додатки та веб-сайти окремих організацій. Чабаботи можна розділити на такі категорії використання, як розмовна торгівля (електронна комерція через чат), аналітика, зв'язок, підтримка клієнтів, дизайн, інструменти для розробників, освіта, розваги, фінанси, харчування, ігри, здоров'я, HR, маркетинг, новини, особисті , продуктивність, торгові, соціальні, спортивні, туристичні та комунальні послуги.

- Соціальний бот

Соціальний бот (також: socialbot або socbot) - це особливий тип chatbot, який використовується в соціальних мережах для автоматичного створення повідомлень (наприклад, твітів) або взагалі виступає за певні ідеї, кампанії підтримки та зв'язки з громадськістю, виступаючи як " послідовник "або навіть підроблений рахунок, який сам збирає послідовників. У цьому відношенні можна сказати, що соціальні боти пройшли тест Тьюринга. Соціальні боти, мабуть, зіграли значну роль на президентських виборах у США 2016 року, і їхня історія, здається, повертається принаймні до середньострокових виборів Сполучених Штатів, 2010 року. За оцінками, 9-15% активних облікових записів Twitter можуть бути соціальними ботами, і 15% загальної кількості користувачів Twitter, які активно обговорювалися на президентських виборчих дільницях, були ботами. Принаймні 400 000 тисяч ботів несуть відповідальність за близько 3,8 млн. Твітів, що складає приблизно 19% від загального обсягу. Всі ці вимоги оскаржуються. Twitterbots - це вже відомі приклади, але відповідні автономні агенти на Facebook і в інших місцях також спостерігаються. Сьогодні соціальні боти можуть створити переконливі інтернет-особи, здатні впливати на реальних людей, хоча вони не завжди надійні. Соціальні боти, крім того, що здатні самостійно створювати повідомлення, також поділяють багато рис з використанням спамбатів у зв'язку з їхньою тенденцією до проникнення великих груп користувачів. Використання соціальних ботів суперечить умовам обслуговування багатьох платформ, зокрема Twitter та Instagram. Проте багато користувачів, особливо підприємства, все ще автоматизують свою діяльність Instagram, щоб отримати справжніх прихильників, а не купувати фальшиві. Це зазвичай здійснюється за допомогою сторонніх компаній соціальної автоматизації. Якщо не будуть прийняті жорсткі правила щодо їх використання, очікується, що соціальні бота грають

важливу роль у майбутньому формуванні громадської думки, автономно виступаючи в ролі безперервних і ніколи не втомливих впливів.

Twitter бот - це тип програмного забезпечення для ботів, який керує обліковим записом Twitter через Twitter API. Програмне забезпечення бота може самостійно виконувати дії, такі як твітинг, повторне твітинг, уподобання, наступні дії, відключення або прямий обмін повідомленнями з іншими обліковими записами. Автоматизація облікових записів Twitter регулюється набором правил автоматизації, які вказують на належне і неналежне використання автоматизації. Належне використання включає в себе передачу корисної інформації, автоматичне створення цікавого або творчого контенту та автоматичне відповідь користувачам за допомогою прямого повідомлення. Неправильне використання включає обхід обмежень швидкості API, порушення конфіденційності користувачів або спаму.

- IRC бот

IRC bot - це набір скриптів або незалежна програма, яка підключається до Internet Relay Chat як клієнта, і тому інші користувачі IRC стають іншим користувачем. IRC-бот відрізняється від звичайного клієнта, а не забезпечує інтерактивний доступ до IRC для користувача-користувача, він виконує автоматичні функції.

- Бот відеоігор

У відеоіграх бот є типовим експертним програмним забезпеченням AI, який відтворює відео ігри на місці людини. Боти використовуються в різних жанрах відеоігор для різних завдань: бот, написаний для шутера від першої особи (FPS), дуже відрізняється від того, що написано для масово багатокористувацької онлайнової рольової гри (MMORPG). Перші можуть містити аналіз карти та навіть базової стратегії; останній може бути використаний для автоматизації повторюваних і втомливих завдань, як сільське господарство. Боти, написані для шутерів від першої особи,

зазвичай намагаються імітувати, як людина грає в гру. Боти, котрі управляються комп'ютером, можуть відтворюватися проти інших ботів та / або гравців у гнідах в унісон, через Інтернет, в локальній мережі або на локальному сеансі. Особливості та інтелект ботів можуть сильно відрізнятися, особливо від вмісту, створеного спільнотою. Розширеними ботами є функція машинного навчання для динамічного вивчення моделей противника, а також динамічне вивчення раніше невідомих карт, тоді як більш тривіальні боти можуть повністю покладатися на списки точок, створених для кожної карти розробником, обмежуючи бот для відтворення лише карт з вказані шляхові точки. Використання ботів, як правило, суперечить правилам поточних масово багатокористувацьких онлайнових рольових ігор (MMORPG), але значна кількість гравців все ще використовує ботів MMORPG для ігор, таких як RuneScape. У MUD, гравці можуть запускати боти для автоматизації складних завдань: ця активність іноді може скласти основну частину геймплея. Хоча заборонена практика в більшості MUD, існує стимул для гравця зберегти свій час, поки бот накопичує ресурси, такі як досвід, для персонажа гравця.

Twitter Bot - це програма, що використовується для створення автоматизованих публікацій, користувачів Twitter або спаму. У цьому проекті я буду використовувати методи машинного навчання, щоб прогнозувати обліковий запис у Twitter, це бот або реальний користувач.

1.2.2 Безпека в соціальних мережах

Люди можуть обмінюватися мультимедійними даними з іншими людьми та підтримувати зв'язок для розваг у соціальних мережах. Щодо користувачів, соціальні мережі подібні до віртуальних комунікаційних середовищ або онлайн-спільноти. Користувачі входять до однієї з цих мереж і шукають нових користувачів з тим самим інтересом після

створення профілю, щоб представити себе. Соціальні мережі демонструють вибуховий ріст за останні роки. Соціальні мережі, такі як Facebook, Twitter та LinkedIn, були дуже популярними і стали переважним методом спілкування для більшості людей. Одночасно популярність соціальних мереж створює значну загрозу для людей. Зловмисники можуть дуже легко отримати важливу особисту інформацію, використовуючи соціальні мережі. Така інформація, така як пароль і банківський рахунок, може допомогти зловмисникам у широкому спектрі мережевих злочинів, включаючи крадіжку особистих даних. Користувачам пропонується на веб-сайтах соціальних мереж вказати ім'я, адресу, стать, дату народження, школу, місце народження, інтерес та іншу особисту інформацію. Цю інформацію буде надано іншим користувачам. Тоді хакери знайдуть важливу інформацію, аналізуючи ці дані. Чим більше користувачів надаватиме інформацію, тим більше зможуть отримати нападники. Деякі соціальні мережі, такі як Twitter, не залишають багато місця для використання, щоб надавати важливу особисту інформацію, але хакери також можуть аналізувати серію цих публікацій і отримати те, що вони хочуть.

Шкідливі програми не є єдиною загрозою. Через необмежений доступ до профілів користувачів, хакери можуть додатково отримати інформацію про корпорацію та комерційні секрети. У дослідженні, проведеному компанією Sophos, це вказує на те, що 1) турбота 62,8 відсотків компаній полягає в тому, що працівники надають забагато інформації в соціальних мережах і 2) 66 відсотків компаній вважають, що використання соціальних мереж стане великою загрозою для корпорацій .

Зловмисники нападають на різні цілі. Ми виявили, що цілі атак у соціальних мережах схожі на Інтернет-сайти. Ці цілі укладаються наступним чином:

- Жарти: Хтось просто хоче пожартувати над іншими користувачами, щоб поліпшити свою репутацію або задовольнити своє власне почуття виконаного обов'язку. Ці атаки не спричинять негативного впливу, але іноді можуть спричинити перевантаженість мережі та змусити користувачів відчувати себе нудно.
- Контроль доступу: Зловмисники контролюють комп'ютери інших користувачів і роблять те, що вони хочуть. Найгірше те, що керовані комп'ютери організовані в Botnet для виконання деяких типів атак, таких як DDOS.
- Особиста інформація: Важлива особиста інформація дуже корисна для нападників. Такі конфіденційності, як пароль, банківський рахунок та номер соціального страхування - це саме те, що нападають шукають. Як тільки хакери отримують цю інформацію, вони можуть вчиняти інші злочини, навіть крадіжку особистих даних.
- Інформація про компанію: в деяких соціальних мережах, таких як користувачі LinkedIn, є бізнес-клієнти. Тож особиста інформація означає величезний резерват багатства. У минулому хакери не могли пробитися через внутрішню мережу, оскільки компанія мала суворі заходи захисту. На відміну від нападників, простіше отримати довіру інших в соціальних мережах. Вони можуть отримати професійну інформацію користувачів та подальшу інформацію про клієнтів. Нарешті, інформація про компанію та інші фінансові секрети піддаються нападам.
- Гроші: ми можемо визнати, що атаки на сайти соціальних мереж стають дедалі більш фінансовими. Крім жартів, найважливішою метою таких атак є гроші. Більшість зловмисників хочуть отримати банківські рахунки, конфіденційність, фінансові таємниці тощо.

Методи виконання атак укладаються наступним чином:

- Спам: розповсюдження безглуздого спаму значно пошкодить доступність мережі. Традиційний спам поширюється по електронній пошті, але зараз вони починають використовувати соціальні мережі. Спам, включаючи рекламу або шкідливий код, може дуже швидко поширюватися за допомогою списку друзів у соціальних мережах.
- Помилка в сторонніх додатках: соціальні мережі, такі як Facebook, дозволяють користувачам додавати сторонні програми, щоб залучити користувачів. Чим більше користувачів додаватиме додаток, тим більше помилок буде принесено. Це призведе до більшої небезпеки.
- Черви : Черв може самовідтворюватись та розповсюджуватись автоматично. Хробак викрадає особисту інформацію, таку як пароль і номер банківського рахунку. Ця інформація буде продаватися на підземному чорному ринку, використовується для крадіжки кредитної картки та банківської інформації користувачів.
- XSS: XSS може бути згенеровано в код веб-сторінки та створює велику загрозу для користувачів. Зловмисники можуть використовувати вразливості XSS, щоб красти COOKIE, викрасти облікові записи, запускати FLASH, змусити користувачів завантажувати шкідливе програмне забезпечення тощо. В соціальних мережах багато взаємодій. Велика кількість інформації, включаючи деякі URL-адреси з виправленням XSS 649 залучить багато користувачів. Після того, як користувачі натискають URL-адресу, напад буде спрацьовувати.
- Плагін: Деякі плагіни, такі як Flash та Silverlight, можуть запускатися в браузері. Це також створює нову загрозу для соціальних мереж. Останнім часом виявлено дефект Flash, і відповідні атаки на соціальні мережі з'являються швидко.

- Фішинг: у соціальних мережах злочинець може приховати себе як законного користувача та використовувати соціальну інженерію, щоб змусити інших користувачів натискати на розроблену URL-адресу. Користувачі в соціальних мережах готові прийняти запрошення незнайомців і спілкуватися з ними. Це призведе до фішингової атаки.

1.2.3 Дезінформація через соціальні ботнети

Ботнети стали головною загрозою в кіберпросторі. Щоб ефективно боротися з ботнетами, ми повинні розуміти командний і контрольний (C & C) ботнету, що є складним завданням, оскільки стратегії та методи C & C швидко розвиваються. Нещодавно бот майстри почали експлуатувати веб-сайти соціальної мережі (наприклад, Twitter.com) як інфраструктуру C & C, що виявиться досить прихованим, оскільки важко відрізнити дії C & C від звичайного трафіку в соціальних мережах. Вони можуть синхронно виконувати команди, що буде призводити до блокування користувачів чи спотворення інформації, дезінформації.

У соціальних мережах, таких як Facebook, Twitter, Instagram боти можуть бути запрограмовані на відправлення скарги на «живого» користувача, що може призвести до його блокування. Механізм блокування такий, що обирається публікація, якій може бути багато років і протягов одної доби на неї скарги за спам чи порнографію. Таким чином соціальна мережа блокує користувача який зробив даний пост на невизначений термін.

1.2.4 Політичні зловживання в соціальних медіа

Приблизно шість років тому технологічні маркетологи використовували соціальні боти для відправлення брутального спаму у формі автоматично поширюваного вмісту соціальних мереж. Зростаюча колекція недавніх досліджень показує, що політичні актори в усьому світі

починають використовувати ці автоматизовані програмні продукти в тонких спробах маніпулювати відносинами та думками в Інтернеті. Політики тепер наслідують популярну тактику для Twitter, що закуповують величезну кількість ботів, щоб значно збільшити кількість прихильників. Міліціонери, державні контрактні фірми та обрані посадові особи використовують політичні боти для розповсюдження пропаганди та інформаційних повідомлень про падіння з політичним спамом.

Політичні боти є одними з останніх і найбільш унікальних технологічних досягнень, розташованих на перехресті політики та цифрової стратегії. Численні інформаційні центри по всьому світу охоплюють розгортання державних та військових бот, приділяючи особливу увагу швидкому зростанню використання такого програмного забезпечення. Журналісти, блогери та громадянські журналісти працювали, щоб пояснити, як уряди та ті, хто борються за владу, використовували програмне забезпечення в певних контекстах. Згідно з повідомленнями ЗМІ, політичні боти були розгорнуті в декількох країнах: Аргентина (Rueda, 2012), Австралія (Peel, 2014), Азербайджан (Pearce, 2013), Бахрейн (Йорк, 2011), Китай (Krebs, 2011), Іран Йорк, 2011), Італія (Фогт, 2012), Мексика (Orcutt, 2012), Марокко (Йорк, 2011), Росія (Кребс, 2011), Південна Корея (Sang Hung, 2013), Саудівська Аравія (Freedom House, 2013) , Туреччина (Poyrazlar, 2014), Сполучене Королівство (Даунес, 2012), Об'єднані держави (Coldeway, 2012) та Венесуела (Ховард, 2014). Нью-Йорк Таймс (Урбіна, 2013) та Нью-Йоркер (Дубін, 2013) опублікували вичерпні статті про появу технології соціального бота, що надають основну інформацію про важливий новий політичний інструмент.

Багато комп'ютерних вчених та розробників політики розглядають трафік, створений через бота, як неприємність, яку можна виявити та управляти. Системні адміністратори на підприємствах, як Twitter, працюють, щоб просто закрити облікові записи, які, як видається,

працюють за допомогою автоматичних скриптів. Ці підходи є надто спрощеними та уникають зосередження уваги на більших та системних проблемах, що виникають у програмному забезпеченні політичного бота. Політичні боти придушують свободу слова та громадянські інновації через демобілізацію активістських груп та задушення демократичної свободи слова. Вони тонко працюють над маніпулюванням громадської думки, надаючи фальшиві враження про популярність кандидатів, сили режиму та міжнародні відносини. Розрив у суспільному житті, викликаний політичними ботами, посилюється новими паралельними обчисленнями та інноваціями до побудови алгоритмів. Тому політичні боти повинні бути краще зрозумілими заради вільної промови та майбутнього цифрової опосередкованої громадської участі. Інформація, яка існує на політичних ботів, не розмежована та часто ізольована до конкретних подій, орієнтованих на країни або на вибори. Цей документ допомагає порівнювати еволюційну траєкторію цього нового середовища, що представляє інтерес у сферах комп'ютерної комунікації, політичного спілкування, інформатики, науки, технологій та суспільства (STS) та інформатики.

Багато досліджень у соціальній науці відображають взаємозв'язок сучасної політики та нових і розвиваються технологій шляхом аналізу звітів ЗМІ про подібні події та інструменти. Оскільки використання політичних ботів є явищем, що виникає, кількість статей англійською мовою, доступних на цю тему, була невеликою.

Country	Year of bot usage
Argentina Australia	2012 2013
Azerbaijan	2012
Bahrain	2011
China	2012
Iran Italy	2011 2012
Mexico	2011
Morocco	2011
Russia	2011
Saudi Arabia	2013
South Korea	2012
Syria	2011
Tibet Turkey	2012 2014
United Kingdom	2012
United States	2011
Venezuela	2012

Рисунок 1.5 - Зареєстровані випадки використання ботів в політиці

На рисунку 5 зображено перші зареестровані випадки дії соціальних ботнетів в деяких країнах. Існує статистика про згуртованість ботів у ботнети а також як автори повідомляють про те, як політичні боти використовуються в різних країнах. Уряди та інші політичні діячі в більшості випадків розгортають політичних ботів під час виборів чи моментів окремої та конкретної країни, політичної бесіди чи кризи. Варто зазначити, що деякі статті також говорили про випадки, коли політичні боти були використані для попередження цілей онлайн-безпеки. Наприклад, сирійський уряд, як повідомляється, використовував боти, щоб створити про регіональну пропаганду, спрямовану як на державні, так і на зовнішні цілі на Twitter під час поточної революції. Описані венесуельські політичні боти зосереджуються виключно на спробах маніпулювати громадською думкою в державі. Кілька журналістів повідомили, що політики в Австралії, Італії, Великобританії та США купували підроблених, ботанізованих, прихильників соціальних медіа в спробах здаватися більш популярними для виборців.

Різний шлях використання політичних ботів залежить від країни до країни та від політичної інстанції до політичної ситуації. Під час виборів політичні боти були використані для демобілізації послідовників протилежної партії. У цьому випадку розгортаючий надсилає Twitter "бомби:" загородження твітів з безлічі бот-керованих облікових записів. Ці твітів кооптують теги, які зазвичай використовують прихильники протилежної партії та повторюють їх твір у тисячі разів, намагаючись запобігти організації серед шахраїв. Наприклад, якщо політичний діяч помічає, що прихильники свого опонента послідовно використовують тег #freedomofspeech в організаційних повідомленнях, тоді цей актор може зробити армію ботів, щоб пролікувати перевизначити цей конкретний тег. Ефект від цього полягає в тому, що прихильникам супротивника дуже складно шукати загальні теги у спробах організувати та спілкуватися зі своїми хлопцями.

Багато випадків використання політичного бота відбувається тоді, коли уряди націлюють на сприйняття загроз кібер-безпеки або політико-культурних загроз з боку інших держав. У деяких статтях згадується державне санкціонування розгортання російського бота. У цих статтях російські боти, як стверджувалося, використовувалися для пропаганди ідеалів режиму або боротьби проти режиму проти цілей за кордоном. Китайські політичні боти нападають на різні країни та комерційні структури в усій Азії та на Заході. Політичні діячі в Азербайджані, Ірані, Марокко, як повідомляється, використовували боти в спробах боротися з антирегіональною мовою та сприяти ідеалам держави.

Уряди, політики та підрядники, які працюють на їхніх роботах, також використовують політичні боти, щоб атакувати державні цілі в соціальних мережах. Описання використання бота в Мексиці є особливою характеристикою цієї автоматизованої стратегії. Згідно з численними джерелами, мексиканський уряд використовував армії Twitter бота, щоб

заглушити публічну інакомислення і ефективно замовчати опозицію через тактику спаму. Peñabots, названий на честь президента Мексики Енріке Пенья Ніето, також використовувався для відправлення проурядової пропаганди. У Туреччині журналісти повідомляють, що уряд президента Реджепа Тайіпа Ердогана та актори від опозиційної республіканської народної армії використовували політичні боти один проти одного, прагнучи як поширювати пропаганду, так і боротися з критикою. У Китаї, а також у китайських адміністративних регіонах Тибету і Тайвані боти були використані для припинення руху суверенітету, одночасно сприяючи ідеалам держави. За словами Кребса, "тибетські співчувачі помітили, що декілька хештейджів Twitter, пов'язаних з конфліктом, включаючи #tibet і #freetibet, зараз так постійно затоплені тюнетами небажаної з явно автоматизованих облікових записів Twitter, що хеш-теги перестали стати корисний спосіб відслідковувати конфлікт ".

Політичні боти були використані під час виборів, щоб надсилати мігрантські повідомлення, які проголосують за кандидатів або є кандидатами. У статті "Нью-Йорк таймс" вказується твердження північнокорейських державних прокурорів про те, що "агенти з Національної розвідувальної служби Південної Кореї в минулому році опублікували понад 1,2 мільйона повідомлень Twitter, щоб намагатися посилити громадську думку на користь Парку Геонг, потім кандидат в президенти та її партія перед виборами 2012 року. "Гуен Х'ю згодом переміг на посаду президента, але начальник розвідки, відповідальний за зусилля, спрямовані на бота, був ув'язнений у в'язницю.

Політичні боти також використовувались під час виборів до списків прихильників соціальних медіа-секретів. У цьому випадку політики заробляють покупців боту, які імітують справжніх користувачів, в спробах виглядати більш популярними або релевантними. Є кілька видатних прикладів, особливо в західних державах. За даними Даунс (2012),

політичний кандидат у Велику Британію Лі Джаспер використовував боти, щоб збільшити кількість своїх послідовників Twitter, щоб "дати помилкове враження про популярність своєї кампанії". Coldeway (2012) деталізує аналогічну пропозицію колишнього президента США кандидат Мітт Ромні, в якому були політичні боти

Висновки до розділу 1

В даному розділі було розглянуто що таке соціальні мережі та які загрози вони несуть. Проаналізувавши загрози та статистику соціальних мереж було обрано цільову (найвпливовішу) соціальну мережу - Твіттер, через те що 340 мільйонів активних користувачів, 92 відсотки політичних лідерів ведуть активну діяльність у цій мережі, зокрема Дональд Трамп, а також через те що в мережі приблизно 15% соціальних ботів. Таким чином останнім фактором для запобігання пропаганди або атак ботнетів залишається аналіз повідомлень (твітів) спеціальним механізмом, робота якого буде полягати в фільтрації стрічки від штучного контенту. Для створення подібного механізму проаналізуємо можливі методи, що дозволяють нам робити припущення про штучність твіта за допомогою аналізу аккаунту користувача, що його створив. Сам аналіз та фінальне припущення робитимемо за ключовими характеристики, за якими зможемо найбільш точно оцінити вхідне повідомлення, завдяки чому матимемо можливість відсіяти штучний контент, що був створений у невідомих цілях.

2 МЕТОДИ МАШИННОГО НАВЧАННЯ ДЛЯ ВИЯВЛЕННЯ БОТІВ

Машинне навчання являє собою науку про надання обчислювальній техніці можливості діяти без явно вказаної інструкції. Замість неї використовується підготовлений набір даних для аналізу, на основі якого будуть прийматися подальші рішення у роботі над новими задачами. За останні роки дана галузь набула широкого застосування у різноманітних сферах, таких як розпізнавання образів, прогнозування, обробки природної мови та інших.

Тобто воно використовується обчислювальних задачах, в яких розробка та програмування явних алгоритмів є нездійсненними. Приклади таких застосувань також включають фільтрацію спаму, оптичне розпізнавання символів (OCR), пошукові системи, виявлення об'єктів та комп'ютерний зір. Машинне навчання іноді поєднують з обробкою даних, де робота фокусується більше на дослідницькому аналізі даних, і є відомою як навчання без учителя.

2.1 Поведінкові шаблони

На даний момент по всьому світу поширюється велика кількість загроз, для яких, в цілях запобігання, служби безпеки направляють підвищену увагу на невербальні способи комунікації, що застосовується в підготовці урядів, військових і правоохоронних органів. З метою забезпечення громадської безпеки потрібне вміння розпізнавати ознаки нестандартної або підозрілої поведінки людини або групи людей, їх поведінкові патерни. Дане розпізнавання може фокусуватися на розумінні рухів, голосових інтонацій, виразів обличчя і жестів [14]. Аналізуються багатоспектральні цифрові фотографії, що демонструють скорочення м'язів обличчя при тих чи інших емоціях, на предмет виявлення патернів емоцій, відчуттів і настроїв.

Новітні системи спостереження використовують програмне забезпечення, здатне відрізняти нормальну поведінку від підозрілої за такими властивості людської поведінки, як інтенсивність, напрямок, форма і патерн. Таким чином комп'ютер відрізняє людей, що ходять, розмовляють, діють нормальним чином від людей ненормальної поведінки, наприклад. Програмне забезпечення, яке використовує нейронні мережі, фіксує і запам'ятовує патерни для створення нових програм, що відрізняють ненормальні патерни за тими ж правилами. Загальне усереднення соціально-поведінкових патернів, що досягається за допомогою вичерпної автоматичної класифікації «нормальності» - в інтересах не тільки широкомасштабних психологічних операцій і технологій політичного контролю, але також і глобального масового маркетингу продуктів споживання.

2.1.1 Виділення ключових характеристик за поведінковими патернами

Як детально описано нижче, передбачено перші три з цих чотирьох функцій показник кількості та якості соціальних взаємодій користувача протягом курсу сеансу. Четвертий (довжина тексту) це замість міри кількості вмісту, виробленого користувачем. Як співвідношення між тривалістю сесії та динамікою показників ефективності спостерігаються в соціальних мережах ми проводимо наш аналіз на сесії аналогічної довжини лише (від 20 до 25 повідомлень), в результаті загалом 1500 сеансів з батом і 13к людських сеансів. Retweet - репортаж з твіту, раніше розміщений іншим користувач Таким чином ми очікуємо збільшення кількості людей retweets під час сеансу, коли користувачі піддаються впливу повідомлення інших користувачів. Частка ретвітів по загальній кількості твітів, згруповані за їхньою позицією в сеансі, показано на малюнку 4а: загалом частка ретвітів вище для людей позиції; користувачі

користувачів збільшують кількість ретвітів на всіх хід їх сесій, починаючи з швидкого зростання в перші 2-3 посади, а потім сповільнюється. Така тенденція не очевидна серед ботів, які здаються замість того, щоб коливатися навколо постійної цінності. Як другий тип взаємодії ми розглядаємо відповідь. Відповідь як випливає з назви, це твіт, розміщений у відповідь на якийсь інший твір Те саме міркування, що і для retweets, застосовуються тут: ми очікуйте, що частка відповідей зростатиме протягом року людські заняття. Наші результати, представлені на рис. 4б, підтверджують наше очікування: як і для ретвітів, частка відповідей зростає і сповільнюється, для людей, за всі перші 20 твітів. Боти, на
з іншого боку, не демонструють аналогічного збільшення. У Twitter користувачі можуть вказати на своїх публікаціях інших користувачів; інший Таким чином, можливим показником соціальних взаємодій є середнє число згадується на посаду. Що стосується попередніх випадків, ми сподіваємось на це число згадується про збільшення, в середньому, в тому, що користувачі входять до системи їх сесії. Результати (мал. 4c) справді свідчать про збільшення середня кількість згадок людей протягом курсу перші 20 твітів. Знову ж таки, боти, здається, не змінюють свою поведінку в Росії хід сесії. Дотепер аналізовані ознаки є всі показники обсягу соціальні взаємодії, в яких залучаються користувачі. Тепер розглянемо середня довжина (у символах) чіріка, що є мірою. Кількість вмісту, що виробляється, і, таким чином, є цікавим показником короткострокова поведінкова динаміка. Перед тим, як порахувати номер з символів, твіт, знятий з усіх URL-адрес, згадок і хеш-тегів, так щоб тільки пояснити частину тексту ефективно складений користувачем. Попереднє дослідження не показало жодних значущих зміна цієї кількості протягом короткострокової сесії на Twitter; проте, аналізи інших платформ показали що середня тривалість посту зменшується на подібних шкалах часу. Тут дані людини показують чітко знижувальну тенденцію, тоді як немає

виникає тенденція до того, що стосується ботів (див. мал. 4d). Зверніть увагу, що для останніх трьох величин (відповідей, згадок і довжина тексту) ми виключили всі retweets з нашого аналізу, як їх вміст не виробляється їхнім плакатом: тоді як сам факт від постановки ретвіт можна вважати поведінковим показником вміст "retweet" навряд чи міг би дати будь-яку корисну інформацію бо як це повага.

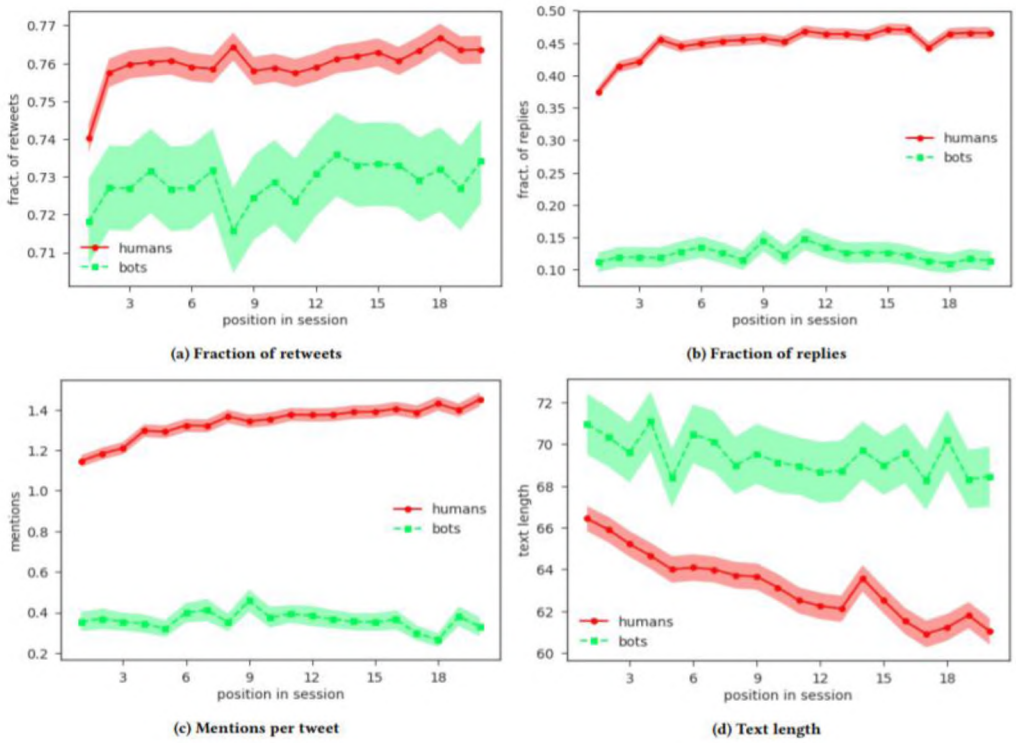

Рисунок 2.1 - Оцінка поведінкових патернів

Тенденція чотирьох різних поведінкових заходів під час онлайн-сесії. Усі сесії що розглядаються тут містити таку ж кількість публікацій (від 20 до 25), щоб обмежити збій через різні моделі поведінки, прийняті користувачами в сесії різної довжини. Кількість розглянутих сеансів, таким чином, становить 1500 для ботів і 13к для людей. З очевидний виняток з рамки 4a, retweets були виключені з нашого аналізу, оскільки нас цікавить тільки оригінал претендувати на виробництво користувачем. Затемнені області відповідають одному SEM, обчисленому окремо для

кожної точки. Для всіх заходів не тільки точки даних добре розділені між двома категоріями користувачів, але люди також показують тимчасові тенденції, які не спостерігаються в ботах. Зокрема: якщо врахувати частку ретвітів (4a), то цінність для користувачів є більшою до ботів на весь курс сеансу; Люди також показують збільшення їх вартості, швидше спочатку (перші 2-3 твітів), потім повільніше, але все ще присутні під час перших 20 твітів. Ситуація аналогічна у випадку відповідей (4b). Людські користувачі також використовуйте більше згадок (4c), з приблизно стійким збільшенням протягом курсу. Прихід до середньої довжини твітів (4d), зменшення виявляється для людей, які також відправляють більш короткі твітів відносно їх автоматизованих аналогів. В всіх чотирьох областях, що розглядаються, не виникає чіткої тенденції для ботів.

2.1.2 Штучно створена природна поведінка

Один з найбільших викликів для виявлення бота в соціальних мережах - це розуміння які сучасні соціальні боти можуть робити. Ранні боти в основному виконувалися один вид діяльності: розміщення вмісту автоматично. Ці боти були наївними та легкими на місці за допомогою тривіальних стратегій виявлення, таких як концентрація уваги на великому обсязі генерації вмісту. У 2011 році команда Джеймса Каврлі в Техаському університеті A & M реалізувала програму. Пастку для honeypot, яким вдалося виявити тисячі соціальних ботів . The Ідея була простою та ефективною: команда створила кілька облікових записів Twitter (ботів), чия роль полягала виключно в створенні безглуздої твітів з непристойним змістом, в якому ніхто не був людиною будь-коли буде цікаво. Однак ці рахунки залучали багато прихильників. Далі інспекція підтвердила, що підозрілі послідовники були насправді соціальними ботами росте їх соціальні кола, сліпо слідкуючи за довільними рахунками. Останнім часом, Twitter боти стають все більш витонченими, роблячи їх

виявлення складніше. Межа між людським і бот-подібним поведінкою є тепер нечіткий. Наприклад, соціальні боти можуть шукати в Інтернеті інформацію та засоби масової інформації заповнюйте їх профілі та надсилайте зібраний матеріал за заданими часами, емулюючи людський тимчасовий підпис змісту виробництва та споживання - в тому числі циркадний закономірності повсякденної діяльності та тимчасові шипи формування інформації. Вони можуть навіть займатися більш складними типами взаємодій, такимияк розважальні бесіди з іншими людьми, коментування їхніх публікацій та відповіді їхні питання . Деякі боти спеціально націлені на досягнення більший вплив, зібравши нових послідовників та розширюючи їх соціальні кола; Вони може шукати соціальну мережу для популярних і впливових людей і дотримуватися їх або зафіксуйте їхню увагу, відправивши їм запити, в надії бути поміченими. Щоб отримати видимість, вони можуть проникнути у популярні дискусії, генеруючилокально-відповідний і навіть потенційно цікавий-зміст, шляхом визначення відповідні ключові слова та пошук в Інтернеті для інформації, яка відповідає цій розмові. Після ідентифікації відповідного вмісту, боти можуть автоматично створювати відповіді за допомогою алгоритмів природної мови, можливо, включаючи посилання на ЗМІ або посилання, що вказують на зовнішні ресурси. Інші боти мають на меті втручання в тотожності законних людей: одні є особистими злодіями, приймаючи невеликі варіанти реальні імена користувачів і крадіжка особистої інформації, такі як зображення та посилання. Навітьможна використовувати більш просунуті механізми; деякі соціальні боти здатні "клонувати"поведінка законних користувачів, взаємодію з друзями та опублікування публікацій когерентного змісту з подібними тимчасовими закономірностями.

2.2 Машинне навчання

Машинне навчання являє собою науку про надання обчислювальній техніці можливості діяти без явно вказаної інструкції. Замість неї використовується підготовлений набір даних для аналізу, на основі якого будуть прийматися подальші рішення у роботі над новими задачами. За останні роки дана галузь набула широкого застосування у різноманітних сферах, таких як розпізнавання образів, прогнозування, обробки природної мови та інших.

Тобто воно використовується обчислювальних задачах, в яких розробка та програмування явних алгоритмів є нездійсненними. Приклади таких застосувань також включають фільтрацію спаму, оптичне розпізнавання символів (OCR), пошукові системи, виявлення об'єктів та комп'ютерний зір. Машинне навчання іноді поєднують з обробкою даних, де робота фокусується більше на дослідницькому аналізі даних, і є відомою як навчання без учителя.

2.2.1 Задачі машинного навчання

Завдання для машинознавства класифікуються у декілька широких категорій. Під керованим навчанням алгоритм будує математичну модель набору даних, яка містить як входи, так і бажані результати. Наприклад, якщо завдання полягає в тому, щоб визначити, чи містить зображення певний об'єкт, дані навчання для керованого алгоритму навчання включатимуть зображення з цим об'єктом або без нього (вхід), і кожне зображення матиме мітку (вихід), що позначає, чи є він містив об'єкт. У деяких випадках вхід може бути частково доступним або обмежений спеціальним зворотнім зв'язком. Напівпроверені алгоритми навчання розробляють математичні моделі з неповних навчальних даних, де частина вхідних зразків відсутня бажаного результату [10].

Алгоритми класифікації та алгоритми регресії - це типи навчального навчання. Алгоритми класифікації використовуються, коли виводи обмежені обмеженим набором значень. Для алгоритму класифікації, який фільтрує електронні листи, вхід буде вхідним електронним листом, а виходом буде назва папки, в якій буде подана електронна адреса. Для алгоритму, який ідентифікує спам-повідомлення, вихід буде прогнозом "спаму" або "не спаму", представленим булевими значеннями 1 і нулем. Алгоритми регресії називаються для їх безперервних виходів, тобто вони можуть мати будь-яке значення в межах діапазону. Прикладами безперервного значення є температура, довжина або ціна об'єкта. У безконтрольному навчанні алгоритм будує математичну модель набору даних, яка містить лише входи та відсутність бажаних результатів. Незабезпечені алгоритми навчання використовуються для пошуку структури даних, наприклад, групування або кластеризації точок даних. Непідтримуване навчання може виявити закономірності в даних і може групувати входи до категорій, як і при вивченні функцій. Зменшення розміру - це процес зменшення кількості "функцій" або входів у наборі даних. Активні алгоритми навчання забезпечують доступ до потрібних результатів (навчальних міток) для обмеженого набору введень на основі бюджету та оптимізації вибору входів, для яких він отримає навчальні етикетки. Коли вони використовуються в інтерактивному режимі, вони можуть бути представлені користувачеві для маркування. Алгоритми навчання зміцнення отримують зворотний зв'язок у вигляді позитивного або негативного підкріплення в динамічному середовищі, і вони використовуються в автономних транспортних засобах або в навчанні грати в гру з протистоянням людини: 3 інші спеціалізовані алгоритми машинного навчання включають тему моделювання, де комп'ютерній програмі надається набір документів з природною мовою та знаходить інші документи, що охоплюють подібні теми. Алгоритми машинного

навчання можуть бути використані для пошуку неперевіреної функції щільності ймовірності в задачах оцінки щільності. Мета алгоритми навчання вивчають власне індукційне упередження на основі попереднього досвіду. В розвиваючій робототехніці алгоритми навчання робота створюють свої власні послідовності навчального досвіду, також відомі як навчальні програми, для кумулятивного набуття нових навичок шляхом самооцінки та соціальної взаємодії з людьми. Ці роботи використовують механізми керування, такі як активне навчання, дозрівання, механізм синергії та імітація.

2.2.2 Методи класифікації у машинному навчанні

Метою роботи є визначення оптимального методу класифікації машинного навчання для аналізу вхідних повідомлень для запобігання витоку інформації, тобто створення механізму, що виявлятиме ймовірність зловмисного використання профілю співрозмовника в цілях здобуття конфіденційної інформації, а саме сповіщення про можливість витоку інформації під час обміну повідомленнями. Таким чином при отриманні нових повідомлень від користувача матимемо певну ймовірність загрози витоку інформації, що базуватимуться на основі попередніх діалогів. Надалі розглянемо деякі з найбільш популярних методів класифікації машинного навчання (k-найближчих сусідів k-NN, метод опорних векторів, баєсів класифікатор, дерево рішень) та проаналізуємо доцільність їх використання у розпізнаванні відправника повідомлення по ключовим характеристикам, що формують поведінкові патерни користувача при веденні переписки у застосунках обміну миттєвими повідомленнями. Також, в наступному розділі, реалізуємо перечисленні методи та порівняємо їх ефективність на спрощеному наборі даних та по результатам роботи виберемо найефективніший класифікатор для вирішення поставленої задачі. Для розв’язання задачі класифікації

власника повідомлення (бот або звичайний користувач), були використані нижченаведені алгоритми машинного навчання.

2.2.2.1 Дерево ухвалення рішень (Decision tree)

Дерево ухвалення рішень (також можуть називатися деревами класифікацій або регресійними деревами) — використовується в галузі статистики та аналізу даних для задач регресії та классифікації. Дерево рішень є схемою, подібною до структури, в якій кожен внутрішній вузол представляє "тест" на атрибуті (наприклад, чи відображається голівкою або хвостом фліп монети), кожна гілка являє собою результат тесту, а кожен аркуш вузол представляє собою label label (рішення приймається після обчислення всіх атрибутів). Шляхи від кореня до листа являють собою правила класифікації.

У процесі аналізу діаграма впливу та рішень дерево рішень використовуються як візуальний та аналітичний інструмент для прийняття рішень, де розраховуються цільова змінна.

Дерево рішень складається з трьох типів вузлів:

- Рішення вузлів - як правило, представлені квадратами
- Шаблонні вузли - типово представлені колами
- Кінцеві вузли - типово представлені трикутниками

Рішення дерев зазвичай використовуються в операціях дослідження та управління операціями. Якщо на практиці рішення повинні бути прийняті в Інтернеті без відкликання за неповними знаннями, дерево рішень має бути паралельним моделлю імовірності як найкращої моделі вибору або алгоритму моделювання в режимі он-лайн вибору. Інше використання дерев рішень є як описовий засіб для обчислення умовних ймовірностей.

Дерева рішень, діаграми впливу, функції корисних функцій та інші інструменти та методи аналізу рішень викладаються студентами старших

курсів у школах бізнесу, економіки охорони здоров'я та охороні здоров'я, а також є прикладами досліджень операцій або методів управління наукою.

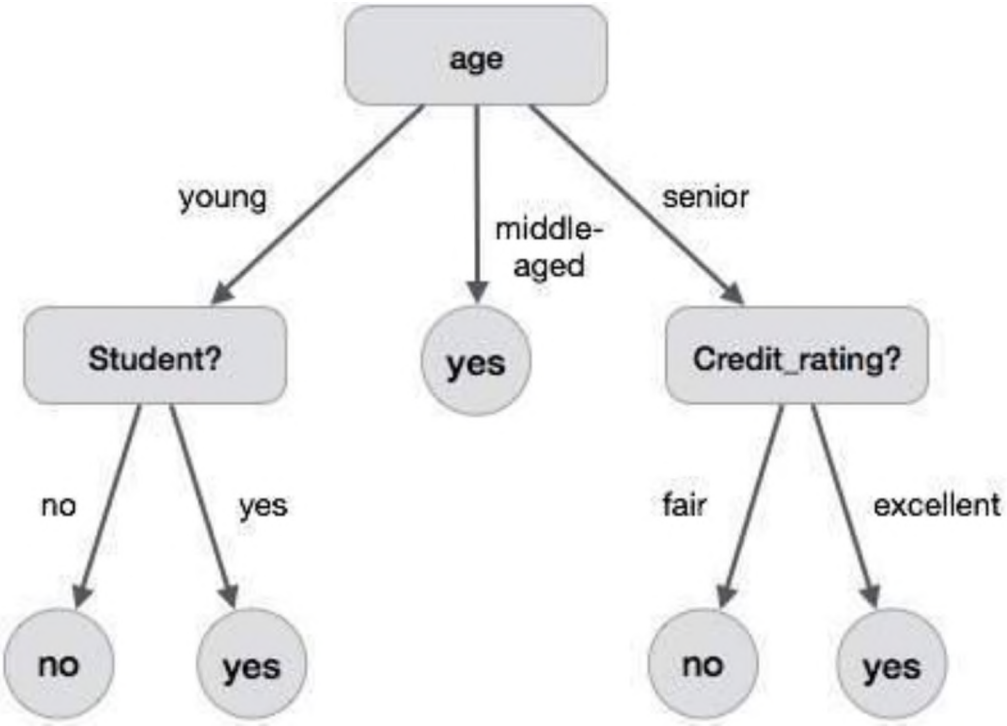

Рисунок 2.2 - Дерево пошуку

Змінна Y є цільовою змінною, тобто тою, яку необхідно класифікувати. Вектор X складається з вхідних змінних X1, X2 , X3 тощо, які використовуються для аналізу та класифікації цільової змінної. Приклад дерева рішень зображено на Рисунку 8.

2.2.2.2 Алгоритм Random Forest

Random forest (з англ. випадковий ліс) є методом вивчення для класифікації, регресії та інших завдань, які працюють шляхом побудови безлічі дерев рішень під час навчання та виведення класу, який є режимом класів (класифікація) або середнім прогнозуванням (регресією) окремих

дерев. Випадкові рішення лісів коректують звичку для прийняття дерев на переобладнанні до їх комплексу тренувань.

Рішення дерев є популярним методом для різних завдань машинного навчання. Вузьке вивчення дерева "наближається до відповідності вимогам, щоб виступати як процедура незавершеного виявлення даних", - кажуть Хастіє та співавтори, - "оскільки вона є інваріантною при масштабування та інших різноманітних перетвореннях значень ознак, є надійною до включення невідповідних функцій і виробляє моделі, що перевіряються, однак вони рідко точні ».

Зокрема, дерева, які вирощуються дуже глибоко, схильні до вивчення дуже нерегулярних моделей: вони накладають набори тренувань, тобто мають низький рівень зсуву, але дуже високу дисперсію. Випадкові ліси є способом усереднення декількох дерев рішень глибоких рішень, навчених на різних ділянках того ж навчального комплексу, з метою зменшення дисперсії. Це відбувається за рахунок невеликого збільшення упередженості та деякі втрати інтерпретації, але в цілому значно підвищує продуктивність у фінальній моделі.

Як частина їх побудови, випадкові прогнозування лісу, звичайно, призводять до неоднорідності серед спостережень. Можна також визначити випадкову різницю між різноманітними лініями між незаміженими даними: ідея полягає в тому, щоб побудувати випадковий прогнозувач лісу, який відрізняє "спостережувані" дані від відповідних синтетичних даних. Дані, що спостерігаються, є оригінальними немаркованими даними, а синтетичні дані наведені з еталонного розподілу. Випадкова несхожість лісу може бути привабливим, оскільки вона обробляє змішані типи змінних дуже добре, інваріантні для монотонних перетворень вхідних змінних і є надійними для зовнішніх спостережень. Випадкова несхожість лісу легко враховує велику кількість напівперервних змінних завдяки своєму власному вибору змінної;

наприклад, "випадкова лісова різниця" "Addcl 1" важить внесок кожної змінної залежно від того, наскільки це залежить від інших змінних. Випадкова несхожість лісу була використана в різних областях застосування, наприклад щоб знайти кластери пацієнтів на основі даних маркера тканини.

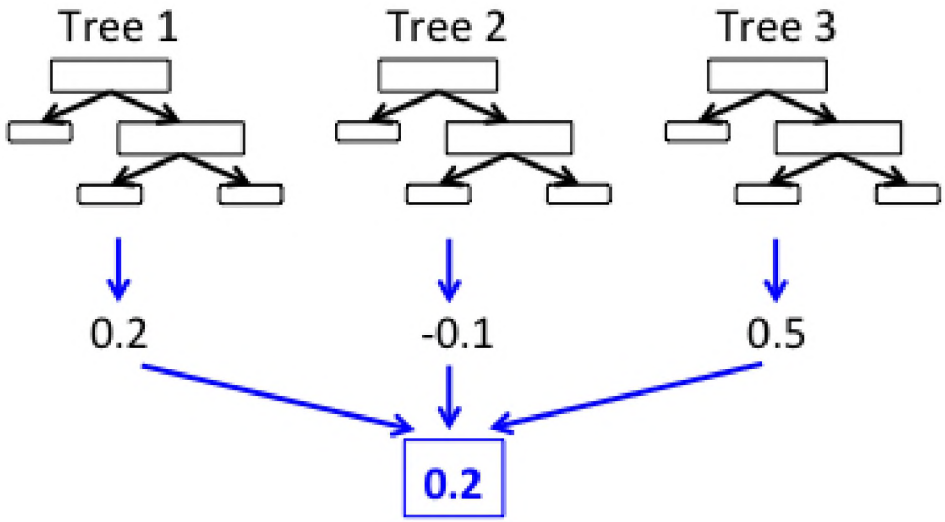

Рисунок 2.3 - Алгоритм Random Forest

2.2.2.3 Наївний баєсів класифікатор

Наївний байєсів класифікатор є сімейством простих "імовірнісних класифікаторів", заснованих на застосуванні теореми Байєса з сильними (наївними) припущеннями про незалежність між ознаками. Наївний Байєс широко поширився 1950-х років. Вона була вперше представлена в спільноту для пошуку тексту на початку і залишається популярним (базовим) методом для класифікації тексту, проблемою судження про документи, що відносяться до однієї чи іншої категорії (наприклад, спам чи законність, спорт або політика і т. д.) з частотою слів, як функції. За умови відповідної попередньої обробки, вона є конкурентоспроможною у цьому домені за допомогою більш просунутих методів, включаючи машини підтримки векторів. Він також знаходить застосування в автоматичній медичній діагностиці.

Класифікатори наївного Байєса мають високу масштабованість, що потребує ряду параметрів, лінійних за кількістю змінних (функцій / предикторів) у проблемі навчання. Тренування максимальної ймовірності можна зробити, оцінюючи вираз із замкнутою формою, який приймає лінійний час, а не дорогою ітераційною наближенням, як це використовується для багатьох інших типів класифікаторів.

У статистиці та комп'ютерній літературі наївні моделі Байєса відомі під різними назвами, в тому числі простими Байєсами та незалежними Байєсами. Всі ці імена вказують на використання теореми Байєса в правилі рішення класифікатора, але наївний Байес не (обов'язково) байєсівський метод.

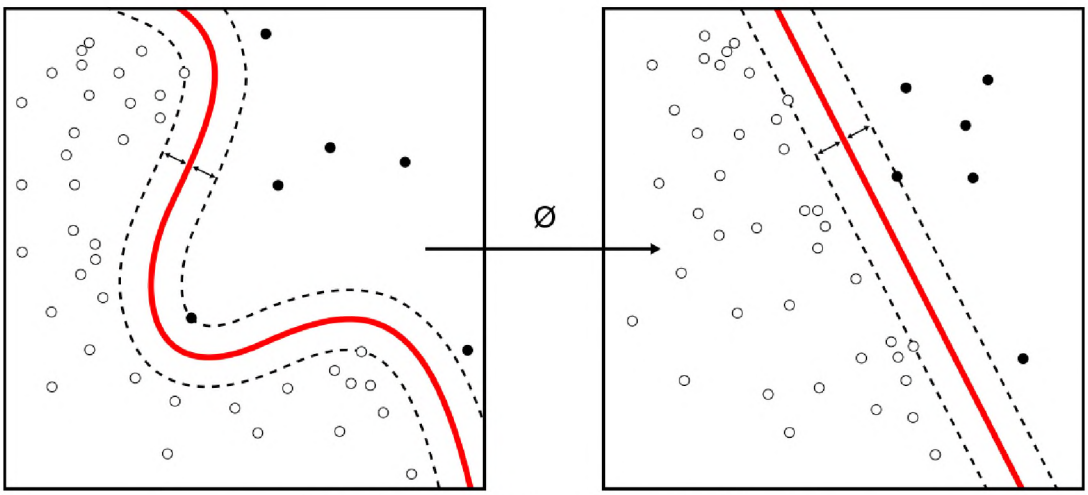

Рисунок 2.4 - Наївний баєсів класифікатор

2.2.2.4 Логістична регресія

У статистиці логістична модель (або модель логіту) є широко використовуваною статистичною моделлю, яка в своїй основній формі використовує логістичну функцію для моделювання бінарної залежної змінної; існує багато складних розширень. В регресійному аналізі логістична регресія (або логітна регресія) полягає в оцінці параметрів логістичної моделі; це форма біноміальної регресії. Математично, бінарна

логістична модель має залежну змінну з двома можливими значеннями, такими як пропуск / невдача, перемога / втрата, жива / мертва або здорова / хворий; вони представлені змінною індикатора, де два значення позначаються як "0" та "1". У логістичній моделі логарифмічні коефіцієнти (логарифм шансів) для значення, визначеного як "1", є лінійною комбінацією однієї або більше незалежних змінних ("прогностичних"); незалежні змінні можуть являти собою двоїнну змінну (два класи, кодовані індикаторною змінною) або неперервну змінну (будь-яке дійсне значення). Відповідна ймовірність значення з позначкою "1" може змінюватися в межах від 0 (звичайно, значення "0") і 1 (звичайно, значення "1"), отже, маркування; функція, яка перетворює log-odds на ймовірність, є логістичною функцією, отже ім'я. Одиниця вимірювання для шкали log-odds називається logit, з логістичного блоку, звідси і альтернативні назви. Також можна використовувати аналогічні моделі з різною сигмоїдною функцією замість логістичної функції, наприклад, модель probit; визначальною характеристикою логістичної моделі є те, що збільшення однієї з незалежних змінних мультиплікативно масштабує шанси даного результату з постійною швидкістю, причому кожна залежна змінна має свій власний параметр; для бінарної незалежної змінної це узагальнює співвідношення шансів. Логістична регресія використовується в різних галузях, включаючи машинне навчання, більшість медичних галузей та соціальні науки. Наприклад, Оцінка тяжкості травми та травми (TRISS), яка широко використовується для прогнозування смертності у постраждалих пацієнтів, спочатку була розроблена Boyd et al. використовуючи логістичну регресію. Багато інших медичних шкал, що використовуються для оцінки тяжкості пацієнта, були розроблені з використанням логістичної регресії. Логістична регресія може бути використана для прогнозування ризику розвитку певної хвороби (наприклад, діабету, ішемічної хвороби серця) на основі спостережуваних

характеристик пацієнта (вік, стать, індекс маси тіла, результати різних аналізів крові тощо). Іншим прикладом може бути прогнозування того, чи індивідуальний виборник голосуватиме на BJP або Trinamool Congress або лівому фронті або конгресі, виходячи з віку, доходу, статі, расової ситуації, стану проживання, голосів на попередніх виборах і т. Д. Техніка може бути також що використовується в машинобудуванні, особливо для прогнозування ймовірності аварії даного процесу, системи або продукту. Він також використовується в маркетингових програмах, таких як прогнозування схильності покупця до покупки продукту або припинення підписки тощо. В економіці це може бути використовується для прогнозування ймовірності обрання людини робочою силою, а бізнес-застосування полягає в тому, щоб передбачити ймовірність того, що власник житла не виконуватиме зобов'язання по іпотеці. У процесі обробки природної мови використовуються умовні довільні поля, розширення логістичної регресії до послідовних даних.

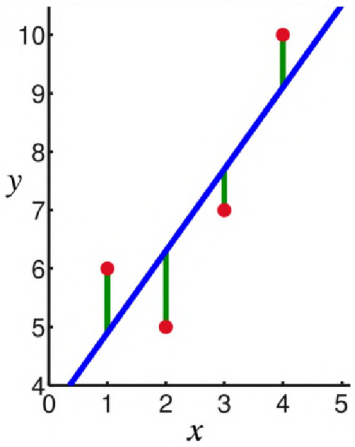

Рисунок 2.5 - Логістична регресія

Пряма, що позначена синім та розділяє площину називається лінійним дискримінантом, так як це пряма то вона має лінійну функцію і дозволяє розподіляти точки на два різних класи. Якщо неможливо провести лінійний розподіл точок у вихідному просторі, перетворюють вектори ознак в простір з великою кількістю вимірювань, додавши додаткові фактори, змінні вищого ступеня і т.д. Використання лінійного

алгоритму в такому просторі дає певні переваги для навчання не лінійної функції, оскільки межа стає нелінійною при поверненні у вихідний простір.

2.2.2.5 Метод k-найближчих сусідів

Метод k-найближчих сусідів (k-NN) є непараметричним методом, який використовується для класифікації та регресії. В обох випадках вхідний файл складається з найближчих прикладів навчання в просторі функцій. Вихід залежить від того, чи використовується k-NN для класифікації чи регресії:

- У класифікації k-NN виходом є членство в класі. Об'єкт класифікується більшістю голосів своїх сусідів, причому об'єкт присвоюється класу, найбільш поширеному серед його найближчих сусідів (k - це позитивне ціле число, зазвичай мале). Якщо k = 1, то об'єкт просто призначається класу того єдиного найближчого сусіда.
- У регресії k-NN виходом є значення властивості об'єкта. Це значення є середнім значенням його найближчих сусідів.

k-NN - це тип навчання на основі екземплярів або лендливе навчання, де функція наближається лише локально, і всі обчислення відкладаються до класифікації. Алгоритм k-NN є одним із найпростіших алгоритмів машинного навчання.

Як для класифікації, так і для регресії, корисна методика може бути використана для присвоєння ваги до внесків сусідів, так що ближчі сусіди становлять більше середнього, ніж більш віддалені. Наприклад, загальна схема зважування полягає у наданні кожному сусідові ваги 1 / d, де d - відстань до сусіда. Сусіди взяті з набору об'єктів, для яких відомий клас (для класифікації kNN) або властивість об'єкта (для k-NN регресії). Це можна розглядати як набір тренувань для алгоритму, хоча ніяких явних

кроків навчання не потрібно. Особливістю алгоритму k-NN є те, що він чутливий до локальної структури даних.

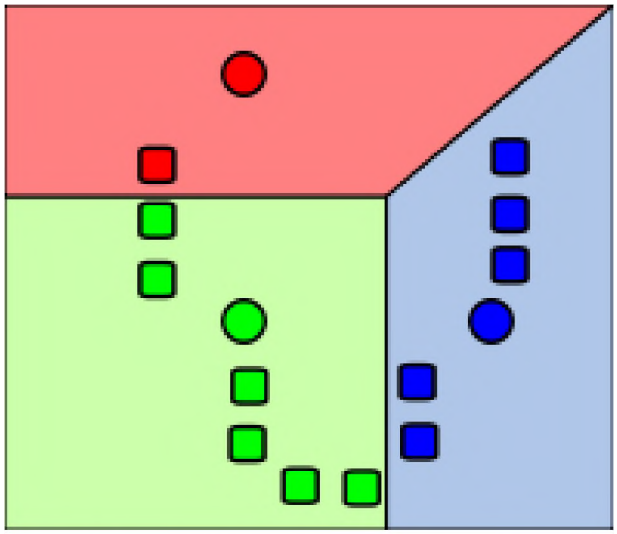

Рисунок 2.6 - Кластеризація.

2.2.2.6 Метод опорних векторів (SVM)

Під час машинного навчання векторні машини підтримки (SVM, також підтримують векторні мережі) контролюються моделями навчання з відповідними алгоритмами навчання, які аналізують дані, що використовуються для класифікації та регресійного аналізу. Враховуючи набір навчальних прикладів, кожен з яких позначається як такий, що належить до тієї чи іншої з двох категорій, алгоритм навчання SVM створює модель, яка призначає нові приклади однієї категорії або іншої, що робить його неімбайвістичним бінарним лінійним класифікатором (хоча методи наприклад, масштабування Platt існує для використання SVM в імовірнісному класифікації). SVM-модель являє собою представлення прикладів як точки в просторі, маповані таким чином, що приклади окремих категорій ділиться явним розривом, який є максимально можливою. Нові приклади потім вказуються в той самий простір і передбачають, що вони належать до категорії, на підставі якої сторони розриву вони падають. Крім виконання лінійної класифікації, SVM можуть ефективно виконувати нелінійну класифікацію, використовуючи

те, що називається ядром, що неявно відображає їхні входи у високорозмірних просторів. Коли дані маркіруються, контрольоване навчання неможливе, і необов'язковий підхід до некерованого навчання, який намагається знайти природне кластеризацію даних групам, а потім картувати нові дані до цих сформованих груп. Алгоритм векторного кластерірування підтримки, створений Хавою Сігельманом та Володимиром Вапніком, застосовує статистику векторів підтримки, розроблених в алгоритмі векторних машин підтримки, для класифікації незамічених даних і є одним з найбільш широко використовуваних алгоритмів кластеризації в промислових цілях.

Приклад роботи класифікатора наведено на Рисунку 2.7

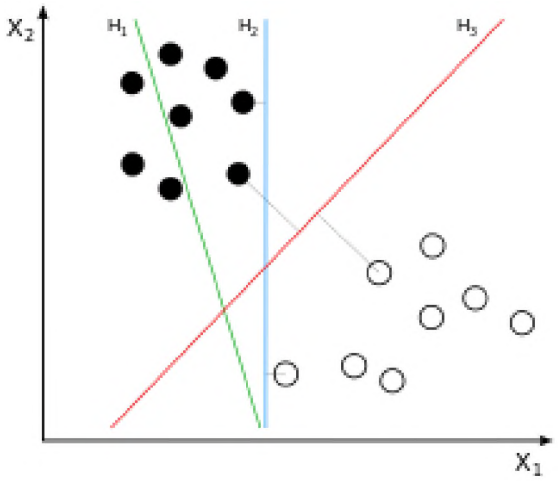

Рисунок 2.7 - SVM

2.2.2.7 Одношаровий перцептон

Перцептрон — алгоритм машинного навчання з учителем для бінарної класифікації, що являє собою функцію що визначає чи належить вхід, представлений числовим вектором, до одного визначеного класу чи ні. Він складається з трьох типів елементів, а саме: сигнали, що надходять від давачів, передаються до асоціативних елементів, а відтак до реагуючих. Таким чином, перцептрони дозволяють створити набір «асоціацій» між вхідними сигналами та необхідною реакцією на виході. Основна ідея

перцептрону заключається в тому, щоб знайти лінійну функцію вектору ознак □(□) = □□□ + □, такій що f(x) > 0 для векторів одного класу та f(x) < 0 для векторів іншого класу, яку називають передавальною функцією. Тут w = w1, w2 ,... , wm є вектором коефіцієнтів(вагою) функції, і b - зсув(англ. bias). Якщо ми позначимо класи чисел через +1 і -1, то нам необхідно буде знайти передавальну функцію вигляду:

□(□) = □□□□(□□□ + □) (2.7)

Данна функція може бути представленою графічно у вигляді нейрону і саме тому перцептрон можна вважати «нейронною мережею». Це найтривіальніше представлення штучної мережі, що містить лише один блок обробки.

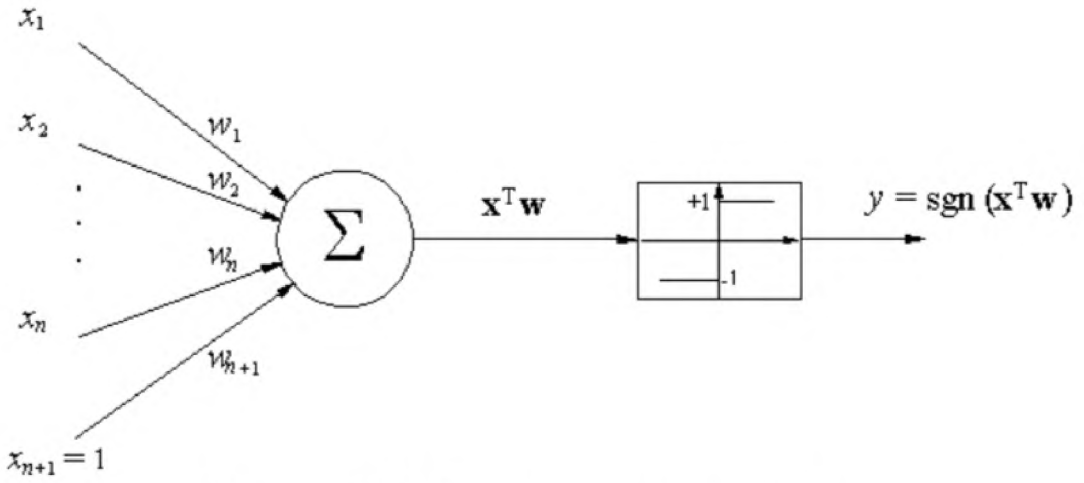

Рисунок 2.8 - Перцептрон у вигляді нейрону

Якщо вектори, що повинні бути класифікованим, складаються лише з двох компонентів (тобто □ ∈ □2), вони можуть бути представлені у вигляді точок на площині. Передавальна функція перцептрона може бути безпосередньо представлена у вигляді лінії, яка ділить площину на дві частини. Вектори в одній півплощині будуть класифікуватися до одного класу, а в інший до іншого класу. Якщо вектори мають 3 компоненти можна, то їх подати в тривимірному просторі, і в загальному випадку n-

розмірні вектори ознак можна подати у вигляді n- розмірного простору. Це підтверджує той факт, що перцептрон є лінійним класифікатором.

Навчання перцептрона здійснюється за допомогою ітеративного алгоритму. Воно починається з довільно обраними параметрами (w0, b0) для передавальної функції та яка ітеративно їх оновлює. На n-ій ітерації алгоритму навчальний приклад (x, c) вибирається таким чином, щоб поточна функція не змогла класифікувати його правильно (тобто □□□□(□□□□ + □□) ≠ □).

Параметри (wn, bn), потім оновлюються за наступними правилами:

$$w_{n+1} = w_n + cx$$

$$b_{n+1} = b_n + c$$

Алгоритм завершує свою роботу, коли передавальна функція встановлює, що правильно класифікує всі навчальні приклади [7]. Даний метод називається методом корекції помилки, тобто вага нейрону не змінюється до тих пір, поки поточна реакція перцептрона залишається правильною. При появі неправильної реакції вага змінюється на одиницю, а знак (+/-) визначається протилежним від знаку помилки. У випадку коли такої функції не існує (тобто поставлена задача не являється лінійно роздільною), алгоритм навчання ніколи не буде збігатися, в такому випадку перцептрон не застосовується.

Якщо дані не лінійно роздільними, то найкраще, що ми можемо зробити - це зупинити алгоритм навчання, коли кількість неправильно класифікованих даних все ще доволі малий. Однак у нашому випадку, данні завжди являються лінійно роздільними, що робить нашу задачу вирішуваною.

2.3 Обробка природних мов (NLP)

2.3.1 Задачі NLP

Обробка природних мов (NLP) - це підполе комп'ютерних наук, інформаційної техніки та штучного інтелекту, що стосуються взаємодії між комп'ютерами та людськими (природними) мовами, зокрема, як програмувати комп'ютери для обробки та аналізу великих обсягів даних природної мови [15].

Виклики в процесі обробки природних мов часто включають розпізнавання мови, природне розуміння мови та створення природної мови. У перші дні багато систем обробки мови були розроблені шляхом ручного кодування набору правил, наприклад написавши граматики або розробляючи евристичні правила для виживання. Проте це рідко залежить від природних змін мови. З часу так званої "статистичної революції" наприкінці 1980-х і середини 1990-х рр. Багато досліджень з обробки природної мови покладалися на машинне навчання.

Парадигма машинного навчання натомість закликає використовувати статистичні висновки для автоматичного вивчення таких правил шляхом аналізу великих корпорацій типових реальних прикладів (корпус (множина, "corpora") - це набір документів, можливо, з анотаціями людини або комп'ютера)

Багато різних класів алгоритмів машинного навчання були застосовані до задач обробки природної мови. Ці алгоритми приймають як вхід великий набір "функцій", які генеруються з вхідних даних. Деякі з найбільш ранніх алгоритмів, наприклад, дерева рішень, випускали системи жорстких правил if-then, подібних до систем рукописних правил, які тоді були загальними. Проте все частіше дослідження зосереджувалося на статистичних моделях, які роблять м'які ймовірнісні рішення, засновані на додаванні реальних ваг до кожної функції вводу. Такі моделі мають перевагу, оскільки вони можуть виражати відносну впевненість у багатьох

різних можливих відповідях, а не лише на одній, що дає більш надійні результати, коли така модель включена як компонент великої системи.

Системи, засновані на алгоритмах машинного навчання, мають багато переваг перед ручними правилами: Процедури навчання, що використовуються під час навчання машин, автоматично зосереджуються на найбільш поширених випадках, тоді як при складанні правил вручну часто не завжди очевидно, де ці зусилля повинні бути спрямовані. Автоматичні методи навчання можуть використовувати алгоритми статистичного висновку для створення моделей, які є надійними для незнайомих введень (наприклад, що містять слова або структури, які раніше не були помічені) та помилковим введенням (наприклад, випадковими словами або словами, які випадково опущені). Як правило, обробка такого входу витончено з правилами, написаними вручну, або, загалом, створення систем рукописних правил, які роблять прийнятні рішення - надзвичайно складно, схильні до помилок та вимагають багато часу.

Системи, що базуються на автоматичному вивченні правил, можна зробити більш точніше, просто надавши більше вхідних даних. Проте системи, засновані на правилах, написаних вручну, можуть бути більш точнішими лише завдяки збільшенню складності правил, що є набагато складнішим завданням. Зокрема, існує межа складності систем на основі ручних правил, за межами яких системи стають все більш і більш керованими. Однак створення додаткових даних для введення в систему машинного навчання просто вимагає відповідного збільшення кількості робочих годин, як правило, без істотного збільшення складності процесу анотації [16].

Методи використання :

1. Синтаксис:

- Граматика індукції - створення формальної граматики, яка описує синтаксис мови)
- Леміматизація - задача полягає у вилученні лише фракційних закінчень і поверненні базової словникової форми слова, що також називається лемою.
- Морфологічна сегментація - задача ділення на окремі слова в спеціальні одиниці мови - морфеми. Складність сильно залежить від складності структури мови. Англійська мова має досить просту морфологію і тому можна просто виділити всі можливі форми слова ("зачиняти, зачинити, зачинено") як незалежні слова. У мовах на кшталт корейської, манджурської, хінді чи японської такий підхід неможливий, тому що існує тисячі можливих форм слова.
- Визначення частини мови у реченні - задача визначення частини мови для кожного слова. Слова, особливо поширені, можуть служити одразу декількома частинами мови. Наприклад, "book" може бути іменником ("book on the table") або дієслово ("book ticket") чт "out" може бути будь-якою з п'яти різних частин мови. Деякі мови не мають таких неоднозначностей, наприклад мови, що мають невелику морфологію, такі як англійська. І навпаки тональні мови під час вербалізації схильні до повної неоднозначності при аналізі написаного тексту.
- Парсинг - граматичний аналіз даного речення. Граматика для природних мов є неоднозначною, а типові речення мають декілька можливих варіантів аналізу.
- Сегментація - це розділення безперервного тексту в окремі слова. Для звичайних мов, наприклад англійської мови, це досить тривіально, оскільки слова, як правило, розділені

пробілами. Тим не менш, деякі письмові мови, такі як китайська, японська та корейська, не містять таких сепараторів, і в цих мовах сегментація тексту є важливою задачею, яка потребує знання словника та деталей морфології мови.

- Пошук термінів - це автоматичне вилучення термінів з деякого тексту.

2. Семантика:
    - Лексична семантика - це визначення сенсу окремих слів у контексті речення.
    - Переклад - це автоматичний переклад тексту з однієї мови на іншу. Це одна з найскладніших проблем в обробці природних мов, бо включає аналіз граматики, семантики, фактів про реальний світ тощо.
    - Визначення власних назв - це визначення, які елементи в тексті належать до власних назв, наприклад людей або місць, а також тип кожного такого імені (наприклад, людина, місцезнаходження, організація). Зауважте, що, хоча велика перша літера може допомогти визнати названі об'єкти в таких мовах, як англійська, це не може допомогти визначити тип іменованого об'єкта і в будь-якому випадку часто є неточним або недостатнім. Наприклад, перша буква речення також велика, а названі об'єкти часто охоплюють кілька слів, лише деякі з яких є капіталізованими. Крім того, багато інших мов взагалі не мають будь-якої великої літери, і навіть мови з великими літерами можуть не використовувати його постійно, щоб відрізняти власні назви. Наприклад, німецька капіталізація всіх іменників, незалежно від того, є вони імена,

а французька та іспанська не використовують імена, які служать прикметниками.

- Природна генерація мови - задача перетворення інформації з комп'ютерних баз даних або семантичних намірів у читабельну людську мову.
- Природне розуміння мови - перетворення шматочків тексту в більш формальні уявлення. Природне розуміння мови передбачає розбір семантики з множинної можливі семантики, яка може бути отримана з природного виразу, який зазвичай набуває форми організованих понять чи речень. Введення та створення мовної моделі та онтології є ефективними, однак, емпіричними рішеннями.
- Оптичне розпізнавання символів (OCR) - це визначення тексту відповідно зображенню, що представляє друкований текст.
- Відповідь на питання - пошук відповіді на питання. Типові запитання мають конкретну правильну відповідь ("Яка столиця України?"), Але іноді також розглядаються відкриті питання ("Коли буде тепло?"). Останні алгоритми розглядають більш складні питання.
- Аналіз настрою (Sentiment Analysis) - задача витягнення суб'єктивної інформації зазвичай з тексту, часто використовуючи ключові слова для визначення "полярності" щодо конкретних об'єктів. Особливо корисно для виявлення тенденцій громадської думки в соціальних мережах з метою маркетингу.
- Сегментація - це задача отримавши частину тексту, відокремити його в сегменти, кожен з яких присвячений окремій темі.

- Словесна розбіжність - багато слів мають більше одного сенсу, ми повинні вибрати слова чи n-грами, які мають найбільшу вагу у речені. Для рішення проблеми зазвичай будують дерево слів і пов'язаних з ними, наприклад з словника або з онлайнового ресурсу, такого як WordNet.

3.Дискурс:

- Автоматична генерація висновків - це генерація головної ідеї тексту та підведення підсумків. Використовується для надання оцінки тексту відомого типу, наприклад, статей по спортивній темі.
- Аналіз текстових посилань - це методика пошуку якорів у реченні тобто які слова "згадки" стосуються тих самих об'єктів. Більш загальне завдання це визначення взаємозв'язків у речені, що включають посилання на вирази. Наприклад, в такому реченні, як "Він увійшов до дому через задні двері", "задні двері" є посиланням, тобто, що двері, про які йде мова, є задніми дверима будинку.
- Аналіз дискурсу - ця визначенні структури дискурсу пов'язаного тексту, тобто характеру діалогу Наприклад коротких відповідей “так” чи “ні”.

2.3.2 Аналіз настроїв

Загалом, аналіз настроїв спрямований на визначення ставлення до спікера, письменника чи іншого суб'єкта щодо певної теми або загальної контекстної полярності або емоційної реакції на документ, взаємодію чи подію. Відношення може бути судженням або оцінкою (див. Теорію оцінки), афективного стану (тобто емоційному стані автора або оратора) або передбачуваного емоційного спілкування (тобто емоційного ефекту, передбаченого автором або співрозмовник). Це завдання зазвичай

визначається як класифікація даного тексту (як правило, речення) в одне з двох класів: об'єктивне або суб'єктивне [17]. Ця проблема іноді може бути складнішою, ніж класифікація полярності. Суб'єктивність слів і фраз може залежати від їх контексту, а об'єктивний документ може містити суб'єктивні речення (наприклад, статті новин, що цитують думку людей). Крім того, як зазначав Су, результати багато в чому залежать від визначення суб'єктивності, що використовується при анотування текстів. Проте Панг показав, що видалення об'єктивних речень з документа перед класифікацією його полярності допомогло підвищити продуктивність. Існуючі підходи до аналізу настроїв можна згрупувати за трьома основними категоріями: технологіями на основі знань, статистичними методами та гібридними підходами. Технології, що базуються на знаннях, класифікують текст за категоріями впливу, що грунтуються на наявності однозначних схожих слів, таких як щасливі, сумні, боязні та нудні. Деякі бази знань не тільки список очевидних впливають на слова, але також привласнюють довільні слова ймовірною "спорідненістю" до певних емоцій. Статистичні методи впливають на елементи машинного навчання, такі як латентний семантичний аналіз, підтримка векторних машин, "мішок слів" та "семантична орієнтація-точка взаємної інформації". Більш складні методи намагаються виявити власника настрою (тобто людини, яка підтримує цей афективний стан) та ціль (тобто об'єкт, про який відчувається вплив). Щоб винести думку в контексті і отримати функцію, про яку говорить оратор, використовуються граматичні відносини слів. Грамматичні відносини залежностей отримуються шляхом глибокого розбору тексту. Гібридний підхід впливає як на машинне навчання, так і на елементи з представлення знань, таких як онтології та семантичні мережі, з метою виявлення семантики, яка виражається тонким чином, наприклад, шляхом аналізу концепцій, які явно не передають відповідну

інформацію, але які неявно пов'язані з іншими концепціями, які роблять це.

Програмні засоби з відкритим вихідним кодом розгортають технології машинного навчання, статистику та методи обробки природної мови для автоматизації аналізу настроїв на великих збірках текстів, включаючи веб-сторінки, онлайн-новини, Інтернет-дискусійні групи, онлайнові огляди, веб-щоденники та соціальні медіа. З іншого боку, системи знань використовують загальнодоступні ресурси, витягуючи семантичну та афективну інформацію, пов'язану з поняттями природної мови. Аналіз настрою також можна виконувати на візуальному вмісті, тобто зображеннях та відео (див. Аналіз мультимодальних настроїв). Одним з перших підходів у цьому напрямку є SentiBank, що використовує прикметник іменованого парного зображення візуального вмісту. Крім того, переважна більшість підходів до класифікації підходів покладаються на модель "мішок", яка ігнорує контекст, граматику та навіть порядок слів. Підходи, які аналізують настрої, засновані на тому, як слова складають значення довгих фраз, показали кращий результат, але вони породжують додаткову додаткову анотацію.

Компонент аналізу людини необхідний для аналізу почуттів, оскільки автоматизовані системи не здатні аналізувати історичні тенденції окремих коментаторів або платформи та часто неправильно класифікуються в їх вираженому настрої. Автоматизація впливає приблизно на 23% коментарів, які правильно класифікуються людьми. Проте люди часто не згодні, і стверджується, що домовленість між людьми забезпечує верхню межу, якою можуть автоматично досягти класифікатори почуттів.

Іноді структура почуттів і тем досить складна. Також проблема аналізу настроїв не монотонна щодо подовження терміну і заміщення зупинки (порівняння їх не дозволять моєму собаці залишатися в цьому

готелі проти я б не дозволив моєму собаці залишитися в цьому готелі). Щоб вирішити це питання, було використано ряд підходів, основаних на правилах та на основі обґрунтування, для аналізу настроїв, у тому числі для неналежного логічного програмування. Також існує декілька правил прокрутки дерева, які застосовуються до дерева синтаксичного аналізу для виявлення актуальності настроїв у налаштуваннях відкритого домену.

- Бізнес: у сфері маркетингу компанії використовують його для розробки своїх стратегій, для розуміння почуттів споживачів щодо продуктів або бренду, як люди реагують на свої кампанії чи продукти, і чому споживачі не купують

продукти

- Політика: у політичній сфері вона використовується для відстеження політичного погляду, виявлення послідовності та невідповідності між заявами та діями на рівні уряду. Це також може бути використано для прогнозування результатів виборів!
- Громадські дії: аналіз настроїв також використовується для моніторингу та аналізу соціальних явищ, для виявлення потенційно небезпечних ситуацій та визначення загального настрою блогосфери.

Висновки до розділу 2

В даному розділі було проведене ознайомилися з основними методами машинного навчання, обробки природного тесту та особливостями поведінкових патернів користувачів або ботів.

Дослідження природи поведінкових патернів та їх специфіки виявило неефективність використання їх для аналізу твітів на штучність. Для побудови моделі прийняття рішення про загрозу повідомлення було проведено аналіз і порівняння наступних методів вирішення задач класифікації: метод найближчих k-сусідів, метод опорних векторів, створення штучної нейронної мережі(перцептрону) та використання наївного баєсового класифікатору. В наступному розділі буде проведено тестування цих моделей для задачі классифікації соціальних ботів та їх активності в соціальній мережі Твітер. На основі принципів обробки природної мови було вирішенно робити аналіз субьективності повідовлення.

3 ПОБУДОВА МЕХАНІЗМУ ДЛЯ ФІЛЬТРАЦІЇ В РЕЖИМІ РЕАЛЬНОГО ЧАСУ

Таблиця 3.1 - Структура аккаунта користувача

Поле	Означення
id	Ціле представлення унікального ідентифікатора для користувача.

id_str	Строкове подання унікального ідентифікатора для користувача.
name	Ім'я користувача, як він його визначив. Не обов'язково ім'я людини. Зазвичай обмежено 20 символами.
screen name	Екранне ім'я, рукоятка або псевдонім, які цей користувач ідентифікує себе. screen_name є унікальними, але можуть бути змінені.
location	Визначене користувачем місце для профілю цього профілю. Не обов'язково розташування, а не машина розбірна. Це поле іноді буде нечітко інтерпретоване службою пошуку.
url	URL, наданий користувачем у зв`язку зі своїм профілем.
description	Користувальницький рядок UTF-8, що описує їх обліковий запис.
verified	Коли це правда, це означає, що користувач має перевірений обліковий запис.

Кінець таблиці 3.1

Поле	Означення
followers count	Кількість підписувачів цього облікового запису в даний час. За певних умов примусу, це поле буде тимчасово позначати "0".
friends count	Кількість користувачів, які використовують цей обліковий запис. За певних умов примусу, це поле буде тимчасово позначати "0".
listed count	Кількість публічних списків, до яких входить цей користувач.

favourites count	Кількість користувачів Tweets, які сподобалося в цьому обліковому записі. Британська орфографія, що використовується в полі назви з історичних причин.
statuses count	Кількість твітів (включаючи ретвітів), виданих користувачем.
created at	Дата UTC datetime, що обліковий запис користувача створено на Twitter.
lang	Код BCP 47 для самодіадезії мови користувача інтерфейсу користувача.
default profile	Якщо це правда, це вказує на те, що користувач не змінює тему або фон свого профілю користувача.
default profile image	Коли це правда, це означає, що користувач не завантажив власне зображення профілю, а замість нього використовується зображення за промовчанням

Надалі ми будемо розглядати сутність з таблиці 2 як основу для класифікації. В наступному підрозділі буде проведена оцінка кореляції цих значень.

3.1 Побудова моделі Твіттер користувача та підготовка набору даних

В попередніх розділах наводилися приклади методів та технологій які використовуються для аналізу та класифікації соціальних ботів, як програм що продукують штучний контент у соціальну мережу Твіттер.

Для навчання алгоритмів та побудови моделей була знайдена вибірка з 2797 користувачів (1321 соціальних ботів та 1476 звичайних користувачів).

Проаналізувавши датасет було виявлено що існують деякі пропущенні значення.

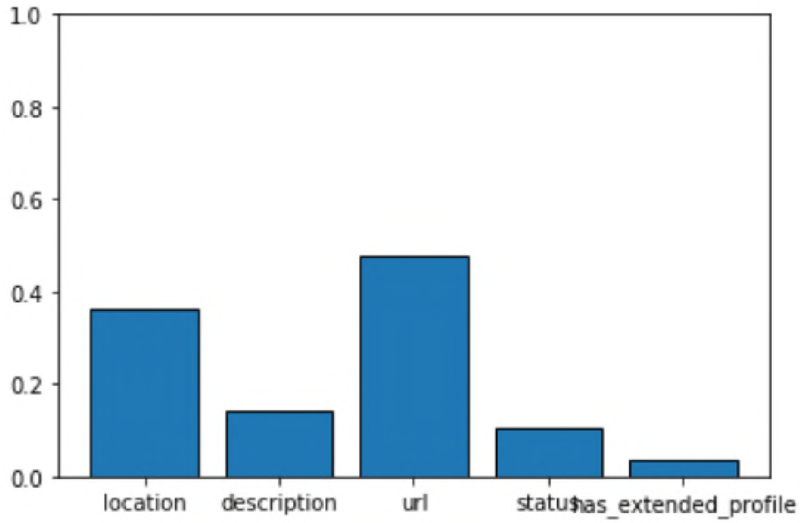

Рисунок 3.1 - Відсутні (Нульові) значення у датасеті

На рисунку 3.1 можна побачити що всі нульові значення припадають на такі колонки:

- location - 38%
- description - 16%
- url - 49%
- status - 9%
- has_extended_profile - 4%

Данні з пропущенними значеннями було вирішено залишити у датасеті через те що данні поля в подальшому не впливають на класифікацію.

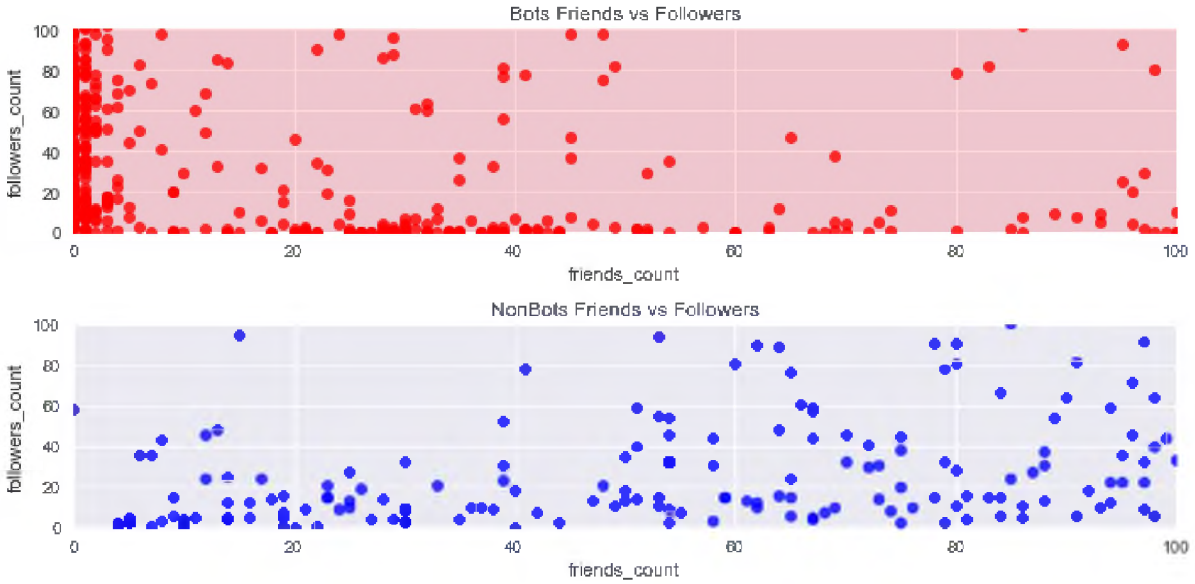

Рисунок 3.2 - Оцінка кореляції кількості підписників і підписок

На рисунку 3.2 було зображено два графіки: залежності кількості фоловерів від кількості друзів для соціальних ботів та реальних користувачів. Як можна побачити на графіках кількість друзів у соціальних ботів у околі 0, так само як і кількість підписок. Для реальних користувачів розподіл зовсім інший - багато друзів та підписок. Виходячи з цього можна використовувати ці дані під час побудови моделі - класифікатора.

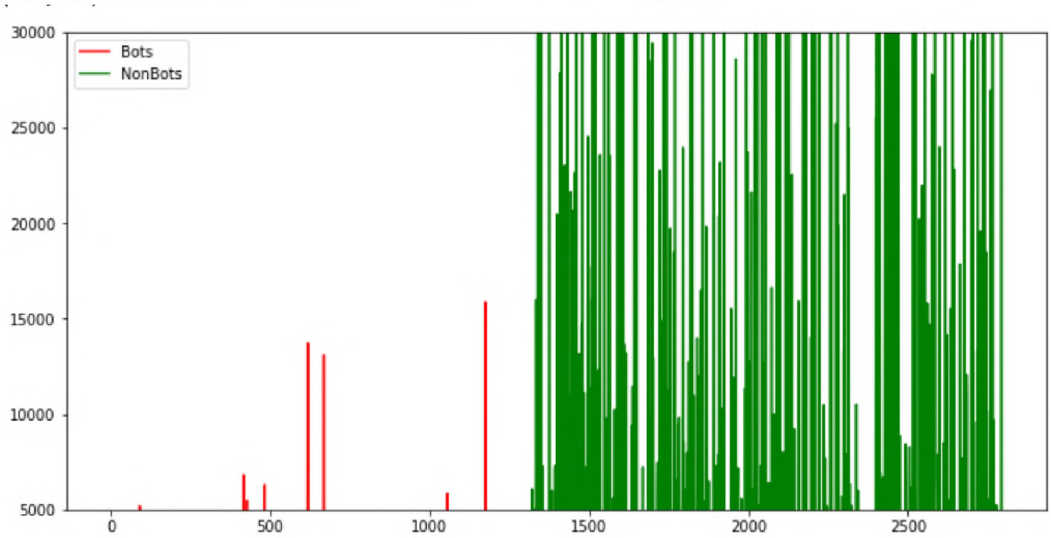

Рисунок 3.3 - Оцінка кількості публічних списків та груп.

На рисунку 3.3 зображено кількість публічних груп та списків у яких знаходиться аккаунт. Проміжок від 5000 до 30000 був обраний для кращої демонстрації кореляції цієї змінної з шуканою. Як видно всі соціальні боти мають значення до 5000 за вийнятком окремих випадків. Для реальних користувачів навпаки всі значення більше 5000, і в більшості випадків більше 30000. Виходячи з цього можна використовувати ці дані під час побудови моделі - класифікатора.

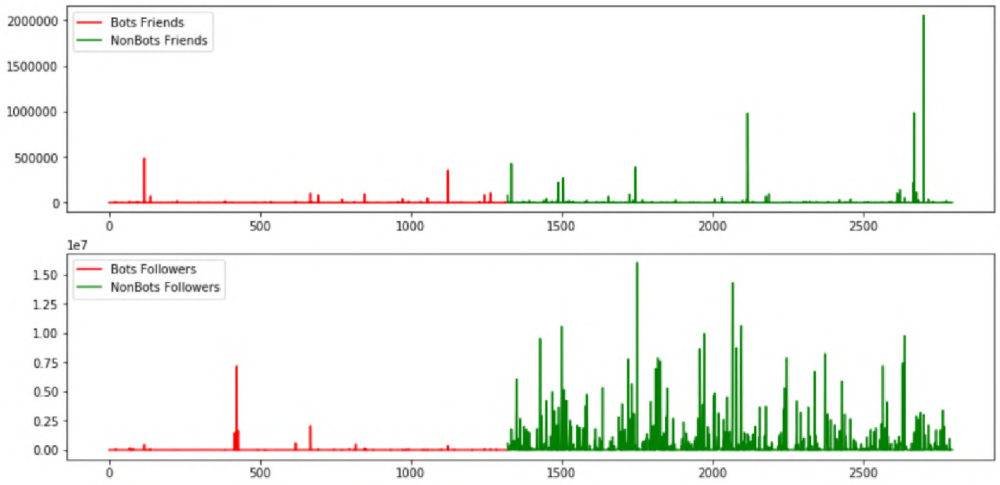

Рисунок 3.4 - Оцінка кількості підписок у соціальних ботів і звичайних користувачів

На рисунку 3.4 зображено кількісну оцінку друзів та підписок для соціальних ботів та реальних користувачів. З графіку можна зробити оцінку залежності кількості підписок від цілбової змінної (бот чи не бот). А також того що у соціальних ботів окрім імітації реальної людьскої поведінки також можуть бути “друзі”. Тобто кількість підписок є більш вагомою величиною ніж кількість друзів.

Проаналізувавши характеристики акаунту можна визначити кореляцію між змінними для пошуку фіч. Спірменовська кореляція між двома змінних дорівнює кореляції Пірсона між ранговими значеннями цих двох змінних; а кореляція Пірсона оцінює лінійні співвідношення, кореляція Спірмена оцінює монотонні зв'язки (будь то лінійний чи ні). Якщо немає повторних значень даних, то ідеальна кореляція Спірмена +1

або -1 відбувається, коли кожна змінна є ідеальною монотонною функцією іншої. Незалежність з використанням кореляції Спірмена:

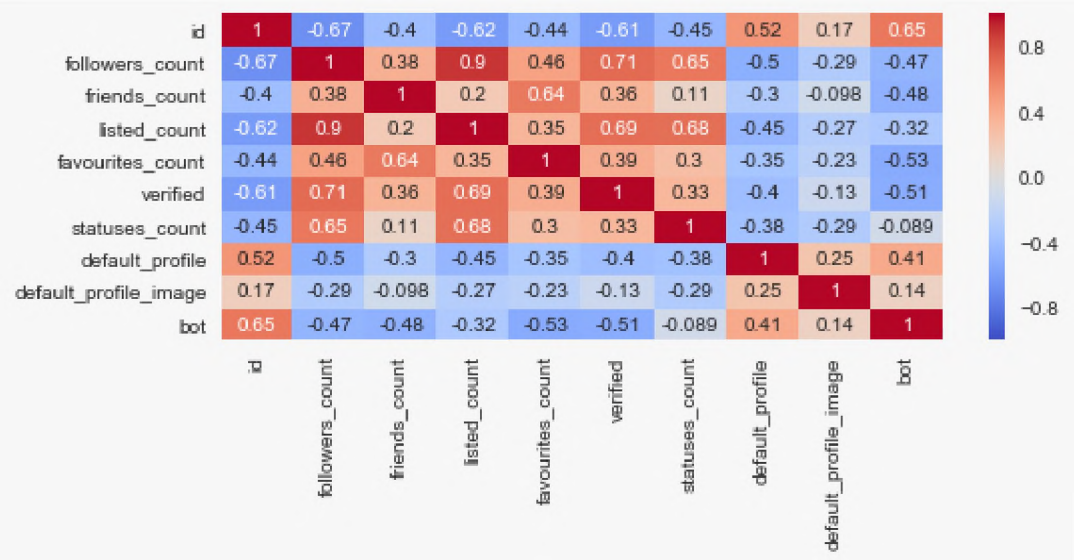

Рисунок 3.5 - Кореляція Спірмена

На рисунку 3.5 зображені всі змінні в аккаунті користувача твіттер та їх коефіцієнт кореляції. Наша цільова змінна - це значення bot ([0,1]). З малюнку можна зазначити що немає ніякої кореляції між id, statuses_count, default_profile, default_profile_image та цільовою змінною (значення в діапазоні [-1,0]), а також існує сильна кореляція між verified, listed_count, friends_count, followers_count та цільовою змінною(значення [0,1]). Найбільш корелюючим та найкращим значенням для побудови моделі є значення listed_count, тож можна зробити його більш вагомим при побудові моделі.

3.2 Вибір та тестування моделі для побудови класифікатора

Розіб'ємо наш датасет у відношенні 3:1 для навчання та тестування відповідно. За частиною для тренування виділимо вектори ознак у бінарні значення, а також перевіримо чи є слово bot або b0t у тестових данних аккаунту. У якості вектору ознак для дослідження виберемо 8 змінних що корелюють з цільовою змінною. Для оцінки результатів роботи алгоритму

класифікації використаємо загальновживані метрики на основі отриманих значень. В залежності від істинного значення тестових даних та результатів передбачення прийнято виділяти ключові поняття:

Tp - true positive — правильне передбачення для позитивного класу

Fp - false positive — хибне передбачення для позитивного класу (помилка першого роду)

Tn - true negative — правильне передбачення для негативного класу

Fn - false negative — хибне передбачення для негативного класу (помилка другого роду)

Різні оцінки будуються на основі різних співвідношень даних значень [13]:

Accuracy(точність) — найпростіша з них, визначається відношенням правильних передбачень до загального розміру вибірки

□□□□□□□□ = □□+□□ / □□ + □□ + □□ + □□

(3.1)

Precision — відображує відсоток вірно передбачених позитивних класів від їх загальної кількості.

□□□□□□□□□ = □□ / □ □ + □□

(3.2)

Recall(повнота) – відображує відсоток вірно передбачених позитивних класів від загального числа класів, визначених моделлю як позитивні.

□□□□□□ = □□ / □ □ + □□

(3.3)

У статистичному аналізі оцінка F1 або F-score є мірою точності тестування. Вона розглядається як точність p, так і повноту тесту, кількість правильних позитивних результатів, поділених на кількість всіх позитивних результатів класифікатору, r - кількість правильних

результатів, поділених на кількість всіх відповідних семплів (всі семпли, які повинні були бути ідентифіковані як позитивні).

(3.4)

Значення досягає найкращого значення в 1 (ідеальна точність і

$$F_1 = \left(\frac{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}}{2} \right)^{-1} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}$$

відкликання), а найгірша - на 0. F-score часто використовується в області пошуку інформації для вимірювання результатів пошуку, класифікації документів та класифікації запитів. F-score широко використовувався в літературі для обробки природної мови, наприклад, оцінці розпізнавання названого об'єкта та сегментації слова.

Таблиця 3.2 - Порівняння моделей

Модель	Tp	Fp	Tn	Fn	Accuracy	Precision	Recall	f1-score
Дерево ухвалення рішень	372	51	64	353	0.93	0.86	0.86	0.86
Random Forest	392	31	62	355	0.94	0.91	0.91	0.91
SVM	360	63	273	144	0.59	0.63	0.6	0.57
Логістична	195	228	22	395	0.7	0.77	0.7	0.68

регресія								
k-найближчих	343	80	61	356	0.85	0.83	0.83	0.83
Перцептон	187	236	22	395	0.68	0.76	0.69	0.67
Наївний баєсів	181	242	12	405	0.69	0.78	0.70	0.67

Після побудови тестових моделей було обрано алгоритм RandomTree для задачі классифікації аккаунту через накйращі показники метричних досліджень (Accuracy, Precision, Recall, f1-score), точність на тестових даних досягла 94% що є найкращим результатом, а також кількість помилок першого і другого роду є не значною.

3.3 Покращення обраної моделі за допомогою ваг

Для покращення обраної моделі потрібно проаналізувати вплив обраних змінних на цільову зміну.

	importance
friends_count	0.361349
followers_count	0.262729
statuses_count	0.163145
verified	0.118208
description_binary	0.051039
status_binary	0.016493
screen_name_binary	0.013551
name_binary	0.011866
listed_count_binary	0.001619

Рисунок 3.6 - Важливість обраних змінних

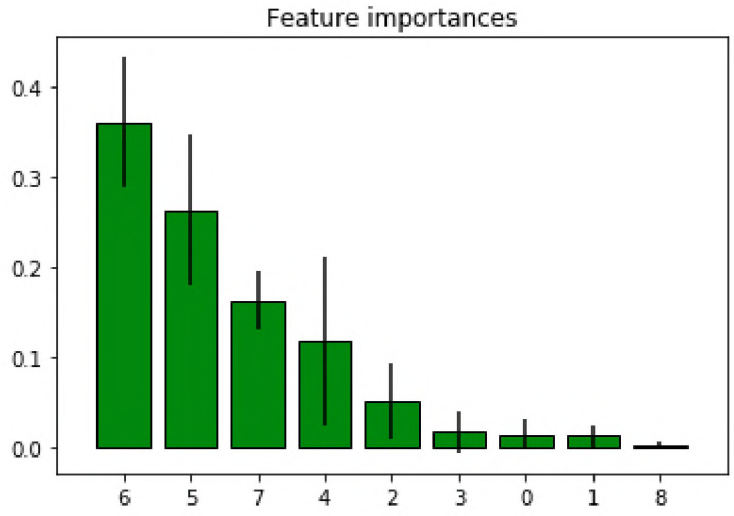

Рисунок 3.7 - Важливість обраних фіч у вигляді таблиці

На рисунках 3.6 та 3.7 можна побачити які значення є більш впливовими на цільову змінну та використати це при побудові фінальної моделі.

3.4 Аналіз суб'єктивності тексту повідомлення

Для аналізу тексту повідомлення у минулому розділі було обрано використання алгоритмів аналізу настроїв (Sentiment Analysis).

В цій сфері є дуже багато рішень і нетренованих моделей. Було обрано реалізацію наївного байесового регресора, а також дані з оглядів на фільми. Вибір був оснований на роботі процесора природних мов TextBlob. Так як не весь контент у соціальній мережі Твіттер англомовний, а приблизно 68%, було вирішено визначати мову твіта перед аналізом, а також перекладати всі не англомовні твіти на англійську мову і оцінювати тільки англомовні варіанти повідомлення. Для визначення мови та перекладу було використано Google Translate API. Алгоритм аналізу настроїв передбачає визначення двох основних параметрів за вхідним

тестом: полярність, суб'єктивність. В даній реалізації їх значення коливаються в діапазоні [-1,1]. Значення менше 0 для полярності означає що речення має негативних характер та відношення до цільових хештегів і навпаки більше 0 означає що речення має позитивне забарвлення, 0 - найтральне, тобто оцінка настрою автора повідомлення відносно цільової теми. Визначення суб'єктивнос це завдання зазвичай визначається як віднесення даного тексту в один з двох класів: суб'єктивний або об'єктивний. Ця проблема іноді може бути більш складною, ніж класифікація полярності: суб'єктивність слів і фраз може залежати від їх контексту, а об'єктивний документ може містити в собі суб'єктивні пропозиції (наприклад, новинна стаття, що цитує думки людей). Тобто 0 значення означає відсутність зв'язаного тексту у реченні, значення більше 0 - суб'єктивність повідомлення, а значення меньше 0 - об'єктивність.

Так як однією з задач соціальних ботів це нав'язування якоїсь думки, то повність суб'єктивні повідомлення можуть бути віднесені у категорії бот чи спам.

3.5 Побудова архітектури фільтра

Додаток для фільтрації був зареєстрований у твіттер як сторонній додаток та використовує access_token, access_secret, consumer_key, consumer_secret для OAuth аутентифікації в системі. Після цього ми підписуємось на користувацькі хештеги які нас цікавлять. Кожного разу як користувач твітера створює повідомлення, воно приходить у фільтр, використовуючи модель для класифікації соціального бота, а також аналіз настрою повідомлення. Повідомлення від ботів відсіюються а також відсіюються повністю субьективні (ті для яких оцінка субьективності дорінює 1).

Після чого, відфільтровані повідомлення відправляються через веб-сокет до клієнтів.

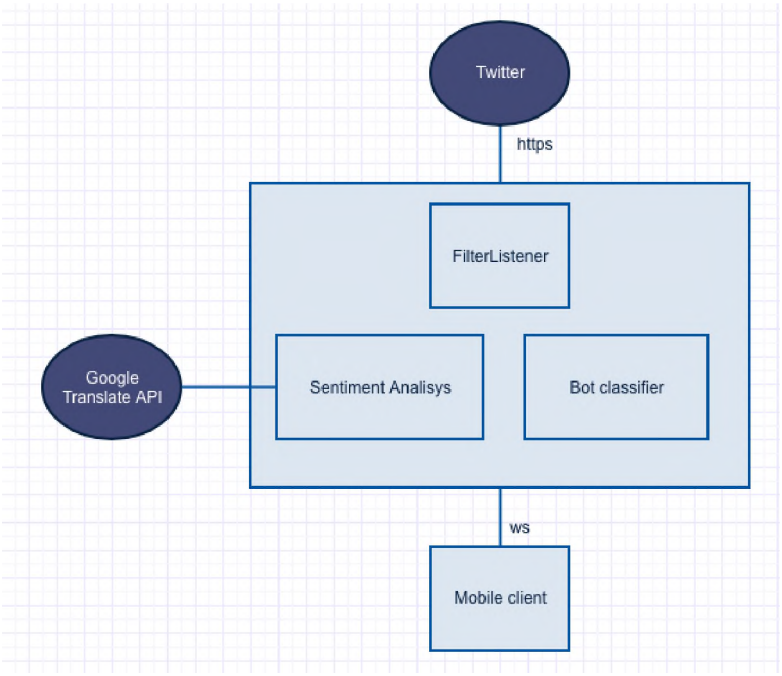

Рисунок 3.8 - Схема додатку.

На рисунку 3.8 зображено схему додатку ws - вебсокет, технологія обміну повідомленнями у режимі реального часу. Під час створення мобільного кліенту було використано кроссплатформовий фреймворк Flutter що дозволяє створювати додатки для обох мобільних платформ (iOS та Android).

3.6 Аналіз результатів

Для аналізу результатів було записано лістинг для аналізу із 841 повідомлення за хештегами java, spring, kotlin, android, ios що класифікувались за допомогою моделі а також оцінювався настрій тексту повідомлення (полярність і суб'єктивність). За результатами роботи було відсіяно 71 твіт. Мануально перевіряючи точність роботи класифікатора було виявлено що серед них було декілька повторень серед авторів, що безсумнівно є ботами (Stack Overflow Jobs, Manning Publications, gocherbot, Android Headlines) і одна помилка, реальний користувач у якого дуже мало друзів.

Рисунок 3.9 - Приклад бота

На рисунку 3.9 показано приклад твіту розробленого ботом, так як тексту в повідомленні немає, а тільки два посилання то обробка тексту повідомлення і є нейтральною (0,0).

Рисунок 3.10 - Приклад реального користувача

На рисунку 3.10 показано приклад твіту розробленого реальним користувачем, текст повідомлення не англомовний, тому його було автоматично перекладено на англійську мову для визначення сенименту, було визначено позитивний настрій користувача відносно теми хештегів та відносну суб'єктивність самого повідомлення.

Висновки до розділу 3

На прикладі тестової вибірки розмічених аккаунтів із соціальної мережі твіттер (соціальних ботів або звичайних користувачів) були проаналізовані дані та виявленні деякі фічі, що корелюють з цільовою зміною, на їх основі було протестовано декілька моделей машинного навчання з параметрами за замовчуванням. Після порівняння результатів було обрано алгоритм Random Forest. Після чого кожна модель була покращена за допомогою параметрів. У підсумку, найкращими алгоритмом для обробки даних про твіттер акаунти виявилась модель побудована на алгоритмі Random Forest, протестувавши даний механізм на реальних даних було встановлено що при записі 841 твіта за хештегами

android, java, spring, kotlin було виявлено 71 твіт від ботів, серед яких лише один був реальним користувачем.

4 РОЗРОБЛЕННЯ СТАРТАП-ПРОЕКТУ

4.1 Опис ідеї проекту

Таблиця 4.1 - Опис ідеї стартап-проекту

Зміст ідеї	*Напрямки застосування*	*Вигоди для користувача*
Впровадження моделі фільтрації даних для Твітер-стрічки від штучного контенту.	1. Розширення для існуючих браузерів (Chrome, Mozilla, Safari, Opera).	Отримання контенту, створеного реальними людьми, без реклами від ботів прямо в браузері.
	2. Додаток для мобільних платформ Android та iOS з вбудованим фільтром.	Отримання контенту, створеного реальними людьми, без реклами від ботів зі свого мобільного пристороюю.

Розробленне рішення не можна порівнювати з іншими, оскільке воно унікальне в своєму роді та не має конкурентів. Основою для нього являється принцип дворівневої фільтрації, тобто оцінки і тексту повідомлення і аккаунту користувача, що його надіслав. Рішення застосовується у ролі додатку в мобільних маркетах чи розширенні для

існуючих браузерів, тому ми можемо лише оцінити його техніко-економічні характеристики, сильні сторони та недоліки.

Таблиця 4.2 - Визначення сильних, слабких та нейтральних характеристик ідеї проекту

№ п/п	Техніко - економічні характеристики ідеї	Мій проект	W (слабка сторона)	N (нейтральна сторона)	S (сильна сторона)
1.	Економічні	Витрати на розробку рішення, закупку ліцензій, розміщення в магазинах додатків, маркетинг - 10000$	Витрати на сервери зі збільшенням користувачів	Немає	Масштабованість, експоненційний ріст користувачів, реферальна система
2.	Технічні	Використання моделі	Портованість між платформами	Немає	Адаптованість до будь-якої платформи, широке коло користувачів
3.	Надійності	Впровадження фільтрування від штучного контенту	Помилки в системі, фільтрація потрібного контенту	Немає	Менше даних проходять через мережу.

4.	Технологічні	Ще один додаток у магазини додатків	Немає	Немає	Немає

Кінець таблиці 4.2

5.	Ергономічні	Підписка на обрані хештеги	Немає	Немає	Менше даних проходять через мережу.
6.	Естетичні	Інтерфейс плагіну або додатку	Немає	Зовнішній вигляд обумовлений інтерфейсом основної платформи	Зручний інтерфейс використання рішення
7.	Екологічність	Обслуговування серверної частини додатку	Немає	Немає	Немає

4.2 Технологічний аудит ідеї проекту

Оскільки наше рішення являється програмним продуктом, основними технологіями будемо розглядати існуючі бібліотеки та SDK для спрощення розробки.

Таблиця 4.3 - Технологічна здійсненність ідеї проекту

№ п/п	*Ідея проекту*	*Технології її реалізації*	*Наявність технологій*	*Доступність технологій*
1.	Побудова плагіну і мобільного кліента	SDK конкретної платформи	Офіційна документація платформи	Доступна
2.	Аналіз тексту	Реалізація баєсового класифікатору мовою Python	Використання бібліотеки TextBlob	Доступна

3.	Аналіз моделі користувача	Реалізація класифікатору RandomForest мовою Python	Використання бібліотек sklearn мови Python	Доступна
Обрана технологія реалізації ідеї проекту: всі технології для реалізації ідеї проекту наявні та доступні				

4.3 Аналіз ринкових можливостей запуску стартап-проекту

Оскільки при плануванні впровадження проекту на ринок ми маємо готові магазини додатків, то основна проблема буде це реклама рішення в цих магазинах.

Таблиця 4.4 - Попередня характеристика потенційного ринку стартап-проекту

№ п/п	*Показники стану ринку (найменування)*	*Характеристика*
1	Кількість головних гравців, од	0
2	Загальний обсяг продаж, грн/ум.од	100 млн ум. од
3	Динаміка ринку (якісна оцінка)	Зростає
4	Наявність обмежень для входу (вказати характер обмежень)	Немає
5	Специфічні вимоги до стандартизації та сертифікації	Немає
6	Середня норма рентабельності в галузі (або по ринку), %	Близько 150%

Таблиця 4.5 - Характеристика потенційних клієнтів стартап-проекту

№ п/п	*Потреба, що формує ринок*	*Цільова аудиторія (цільові сегменти ринку)*	*Відмінності у поведінці різних потенційних цільових груп клієнтів*	*Вимоги споживачів до товару*
1	Велика кількість спаму та ботів у соціальній мережі Твітер.	Будь який користувач мережі інтернет.	Основна цільова група користувачів продукту - ділові партнери, бізнесмени, люди що володіють потенційно важливою інформацією. Друга група - прості користувачі, які бажають зменшити рівень спаму	- фільтрація даних від спаму та штучного контенту.

Таблиця 4.6 - Фактори загроз

№ п/п	*Фактор*	*Зміст загрози*	*Можлива реакція компанії*
1	Недовіра користувачів до нового додатку	Зменшення кількості користувачів визване порушеннями конфіденційності інформації або її використання тощо	Відсутність впливу на існуючі рішення призведе до потреби інтеграції з рішеннями від Твітеру.
2	Можливість не побачити цікаву рекламу чи статистику для користувача	Надлишкова фільтрація даних	Змінення критеріїв фільтрації

Таблиця 4.7 - Фактори можливостей

№ п/п	*Фактор*	*Зміст можливості*	*Можлива реакція компанії*
1	Почастішання випадків спаму чи агітації в інтернеті.	Зростаюча потреба користувачів в надійному способі попередити подібні загрози	Розширення клієнтської бази, маркетингові дії
2	Покращення рівня підтримки додаткових інтеграцій з соціальною мережею Твітер	Сприяння компаній систем фільтрації та розробка нових рішень	Збільшення масштабів розвитку

Таблиця 4.8 - Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	*В чому проявляється дана характеристика*	*Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)*
1. Вказати тип конкуренції: Монополія	Заборона або суттєве обмеження повноважень додатків, неможливість продовження відповідного функціонування	Необхідність прийняття нових правил роботи або створення власного рішення
2. За рівнем конкурентної боротьби: Глобальний	Витіснення або обмеження компаніями додатків з своїх систем	Створення власного рішення
3. За галузевою ознакою Внутрішньогалузева	Рішення застосовується в одній галузі	Розширення сфери надання послуг та функціоналу
4. Конкуренція за видами товарів: - товарно-видова	Рішення використовується для задоволення потреб клієнтів, але істотно відрізняються від рішень конкурентів на користь якості існуючих сервісів	Впровадження рішення в дрібні системи обміну повідомленнями, корпоративні чати
5. За характером конкурентних переваг - нецінова	Вартість не є ключовим фактором для клієнтів	Покращення якості продукту та збільшення функціоналу збільшує клієнтську базу
6. За інтенсивністю - марочна	Результуюча привабливість продукту	Підвищення якості роботи механізму

Таблиця 4.9 - Аналіз конкуренції в галузі за М. Портером

	Прямі конкуренти в галузі	*Потенційні конкуренти*	*Постачальники*	*Клієнти*	*Товари замінники*
Складові аналізу	*Навести перелік прямих конкурентів: Відсутні*	*Визначити бар'єри входження в ринок: Botometer*	*Визначити фактори сили постачальників: Не впливають*	*Визначити фактори сили споживачів: Відмова від користування продуктом або ж навпаки визнання*	*Фактори загроз з боку замінників: Повернення до листування, email, дзвінків*
Висновки	Визначити інтенсивність конкурентної боротьби з боку прямих конкурентів: Відсутня	Ріщення допогая тільки дізнаватися чи є користувач Твітера ботом чи ні.	Чи постачальники диктують умови роботи на ринку? Які? Не диктують умови	Чи клієнти диктують умови роботи на ринку? Які?: Вимоги до якості продукту, точності його роботи	Обмеження для роботи на ринку через товари замінники: Не передбачається

Таблиця 4.10 - Обгрунтування факторів конкурентоспроможності

№ п/п	*Фактор конкурентоспроможності*	*Обгрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)*
1	Відсутність прямих конкурентів	Унікальність розробленого рішення, використання декількох методів фільтрації
2	Багатоплатформеність	Впровадження моделі в усі існуючі платформи.
3	Доступність	Простота в використанні клієнтами, велика кількість користувачів

Кінець таблиці 4.10

4	Зручність	Зручний інтерфейс у використанні, обумовлений інтерфейсом основної платформи
5	Багатофункціональність	Застосування підходить для різної цільової аудиторії

Порівняльний аналіз сильних та слабких сторін “CleanTwitter”.

Оскільки прямих конкурентів немає, ми не можемо визначити відносний рейтинг нашого рішення серед інших. Зазначимо лише, що воно надасть додатковий рівень захищеності при спілкуванні шляхом автентифікації користувачів та їх повідомлень.

Таблиця 4.11 - SWOT-аналіз стартап-проекту

Сильні сторони: - практична корисність продукту - відсутність конкурентів - незалежність від політично економічного стану країни	Слабкі сторони: - залежність від політики компаній Twitter - потреба у використанні користувачами сторонніх додтаків
Можливості: - підтримка різних платформ - збільшення інтересу користувачів до фільтрації їх Твіттер стрічки через ріст кількості спам-ботів.	Загрози: - обмеження функціональності платформами Twitter - втрата довіри та вихід Twitter з ринку.

Таблиця 4.12 - Альтернативи ринкового впровадження стартап-проекту

№ п/п	*Альтернатива (орієнтовний комплекс заходів) ринкової поведінки*	*Ймовірність отримання ресурсів*	*Строки реалізації*
1	Повноцінна реалізація для однієї платформи	Дуже висока	2-3 місяця
2	Підключення до основних платформ	Висока	4-7 місяців
3	Розробка інтегрованого у Твітер рішення	Ймовірна, при домовленості з компанією	1 рік

Отже, з перелічених альтернатив, найкращим рішенням буде початок роботи з повноцінної реалізації продукту для однієї платформи, та поступовим розширенням і переходом до багатьох платформ.

4.4 Розроблення ринкової стратегії проекту

Таблиця 4.13 - Вибір цільових груп потенційних споживачів

№ п/п	*Опис профілю цільової групи потенційних клієнтів*	*Готовність споживачів сприйняти продукт*	*Орієнтовний попит в межах цільової групи (сегменту)*	*Інтенсивність конкуренції в сегменті*	*Простота входу у сегмент*
1	Звичайні користувачі	Користувачі зацікавлені в збереженні свого часу	Середній, попит у користувачів що обмінюються інформацією	Відсутня	Простий вхід, необхідність проведення маркетингових компаній

Кінець таблиці 4.13

2	Організатори івентів	Компанії, які використовують Твітер для своєї роботи	Середній	Відсутня	Можливі складнощі з зміною політики компаній щодо впровадження продукту
Які цільові групи обрано: звичайні користувачі та організатори івентів.					

За результатами аналізу потенційних груп споживачів будемо використовувати стратегію диференційованого маркетингу.

Таблиця 4.14 - Визначення базової стратегії розвитку

№ п/п	*Обрана альтернатива розвитку проекту*	*Стратегія охоплення ринку*	*Ключові конкурентоспроможні позиції відповідно до обраної альтернативи*	*Базова стратегія розвитку*
1	Повноцінна реалізація для однієї платформи	Стратегія диференційованого маркетингу	Підвищення рівня захищеності обміну повідомленнями	Стратегія диференціації
2	Підключення до основних платформ	Стратегія диференційованого маркетингу	Розширення ринку продукту	Стратегія диференціації
3	Розробка інтеграції з соціальної мережею Твітер	Стратегія диференційованого маркетингу	Виділення власного продукту з усіма перевагами у використанні	Стратегія спеціалізації

Таблиця 4.15 - Визначення базової стратегії конкурентної поведінки

№ п/п	*Чи є проект «першопроходцем» на ринку?*	*Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?*	*Чи буде компанія копіювати основні характеристики товару конкурента, і які?*	*Стратегія конкурентної поведінки*
1	Так	Рішення буде доступне всім користувачам існуючих платформ	Ні	Стратегія лідера, розширення первинного попиту

Таблиця 4.16 - Визначення стратегії позиціонування

№ п/п	*Вимоги до товару цільової аудиторії*	*Базова стратегія розвитку*	*Ключові конкурентоспроможні позиції власного стартап-проекту*	*Вибір асоціацій, які мають сформувати комплексну позицію власного проекту (три ключових)*
1	Фільтрація контенту від спаму та реклами	Стратегія диференціації	Фільтрація контенту від спаму та реклами	Фільтрація контенту від спаму і реклами Використання на всіх платформах Безпечне використання Твіттеру

4.5. Розроблення маркетингової програми стартап-проекту

Таблиця 4.17 - Опис трьох рівнів моделі товару

Рівні товару	*Сутність та складові*		
I. Товар за задумом	Забезпечення фільтрації даних від штучного контенту.		
II. Товар у реальному виконанні	Властивості/характеристики	М/Нм	Вр/Тх /Тл/Е/Ор
	1. Використання моделі користувача	М	Тх, Е
	2. Аналіз тексту	М	Тх, Е
	3. Робота на всіх платформах	М	Тх, Тл, Е
	4. Інтеграція з соціальною мережою	М	Тл, Е
	Якість: RFC 6238(TOTP), RFC 4226 (HOTP)		
	Марка: CleanTwitter		
III. Товар із підкріпленням	До продажу: використання безкоштовної версії з обмеженим функціоналом		
	Після продажу: надання послуг без обмежень, проведення додаткового маркетингу		
За рахунок чого потенційний товар буде захищено від копіювання: захист інтелектуальної власності, унікальний підхід в фільтрації даних.			

Таблиця 4.18 - Визначення ключових переваг концепції потенційного товару

№ п/п	*Потреба*	*Вигода, яку пропонує товар*	*Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити*
1	Фільтрація контенту від спаму та реклами	Впровадження додаткового рівня фільтрації контенту із Твіттер стрічки	Автоматичне рішення

Таблиця 4.19 - Визначення меж встановлення ціни

№ п/п	*Рівень цін на товари-замінники*	*Рівень цін на товари-аналоги*	*Рівень доходів цільової групи споживачів*	*Верхня та нижня межі встановлення ціни на товар/послугу*
1	Відсутній	Відсутній	Будь-який	Для клієнтів 5у.о. на місяць; для компаній 40/300/500у.о на місяць

Таблиця 4.20 - Формування системи збуту

№ п/п	*Специфіка закупівельної поведінки цільових клієнтів*	*Функції збуту, які має виконувати постачальник товару*	*Глибина каналу збуту*	*Оптимальна система збуту*
1	Пошук методів фільтрації від спаму	Підтримка користувачів, оновлення програмного рішення	Магазини додатків та розширень	Продажі через магазин

Таблиця 4.21 - Концепція маркетингових комунікацій

№ п/п	*Специфіка поведінки цільових клієнтів*	*Канали комунікацій цільових клієнтів*	*Ключові позиції, обрані для позиціонування*	*Завдання рекламного повідомлення*	*Концепція рекламного звернення*
1	Тестування продукту	Твіттер та інші соціальні мережі	Фільтрація Твітер-стрічки від спаму та активності ботів.	Вказати на можливі та набридливі повідомлення від ботів	Демонстрація можливих та набридливі повідомлення від ботів

Висновки до розділу 4

В результаті проведеного аналізу бачимо, що проект має непогані можливості ринкової комерціалізації. Має перспективи впровадження з огляду на потенційні групи клієнтів, а саме користувачів соціальної мережі Твіттер. Бар'єри входження полягають лише в готовності користувачами використовувати дане рішення замість звичайної Твіттер-стрічки Конкуренція для нашого рішення відсутня.

Для впровадження нашого рішення, найкраще почати з інтеграції з однією з розширенням для браузеру, та поступово проводити інтеграції з іншими системами. В перспективі, при зростанні клієнтської бази можливе виділення в окрему систему, тому подальша імплементація проекту є доцільною.

ВИСНОВКИ

Результатом виконання дипломної роботи є модель фільтрації даних та програмний інструммент побудований на її основі. У процесі аналізу та побудови моделі були виконані наступні завдання :

- Проаналізовано існуючі соціальні мережі, обрано найвпливовішу (Твітер).
- Проаналізовано існуючі методи пошуку контенту, створеного ботами.
- В результаті аналізу було обрано фільтрувати потік повідомлень на основі інформації про аккаунт, що створив повідомлення та тексту повідомлення.
- Проведено аналіз датасету акаунтів та виявлено кореляцію між даними.
- Проведено аналіз алгоритмів класифікації на обраному датасеті та побудовано модель користувача.
- За допомогою обраних факторів та моделей машиного навчання побудовано модель та впровадженно в інструмент який дозволяє здійснювати фільтрацію контенту в режимі реального часу.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1 Social Network Analysis in the Social and Behavioral Sciences [Текст] / Wasserman Stanley, Faust Katherine, 1994. 1-27 с.

2 Network Analysis in the Social Sciences [Текст] / Borgatti Stephen P.; Mehra Ajay; Brass, Daniel J., Labianca Giuseppe, 2009. 892–895 с.

3 The Advantages And Disadvantages Of Social Networks [Електронний ресурс] – 2013. – Режим доступу до ресурсу: https://www.ukessays.com/essays/internet/advantages-and-disadvantages-of-social-networks.php

4 Number of social network users worldwide [Електронний ресурс] – 2018. – Режим доступу до ресурсу: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

5 Social Media Active Users by Network [Електронний ресурс] – 2018. – Режим доступу до ресурсу: https://www.thesocialmediahat.com/active-users

6 Amazing Social Media Statistics and Facts [Електронний ресурс] – 2018. – Режим доступу до ресурсу: https://www.brandwatch.com/blog/amazing-social-media-statistics-and-facts/

7 Most famous social network sites worldwide as of October 2018 [Електронний ресурс] – 2018. – Режим доступу до ресурсу: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

8 The Rise of Social Bots [Текст] / Emilio Ferrara, Onur Varol, Clayton A. Davis, Filippo Menczer, Alessandro Flammini, 2017. 371–411 с.

9 Feature Engineering for Machine Learning and Data Analytics [Текст] / Guozhu Dong, Huan Liu, 2017. – 558 с

10 Machine Learning Algorithms [Текст] / Giuseppe Bonaccorso, 2017. – 234-245 с

11 Practical Statistics for Data Scientists [Текст] / Andrew Bruce, Peter Bruce, 2016. – 245 с

12 An experimental comparison of naïve Bayesian and keyword-based anti-spam filtering with personal e-mail messages Proceeding of the an International ACM SIGIR Conference on Res and Devel in Inform Retrieval [Текст] / I. Anderouysopoulos, J. Koutsias, K.V. Chandrianos, G. Paliouras, C. Spyrpolous. – 2000. – 465 с.

13 Metrics To Evaluate Machine Learning Algorithms in Python [Електронний ресурс]/ Jason Brownlee – 2016. – Режим доступу до ресурсу: https://machinelearningmastery.com/metrics-evaluate-machine-learning-algorithms-python/

14 Measuring bot and human behavioral dynamics [Електронний ресурс]/Iacopo Pozzana, Emilio Ferrara – 2018. – Режим доступу до ресурсу: https://arxiv.org/pdf/1802.04286.pdf

15 Exploring the Limits of Language Modeling [Електронний ресурс]/ Rafal Jozefowicz, Oriol Vinyals, Mike Schuster, Noam Shazeer, and Yonghui Wu - 2016.- Режим доступу до ресурсу: https://arxiv.org/abs/1602.02410

16 Natural Language Processing with Python. [Електронний ресурс]/ Steven Bird, Ewan Klein, and Edward Loper - 2009.- Режим доступу до ресурсу: O'Reilly Media. ISBN 978-0-596-51649-9.

17 Opinion Mining from Noisy Text Data [Текст] /Dey, Lipika; Haque, S. K. Mirajul - 2008. - с.83-90.

ДОДАТКИ

Додаток А

Використання моделі класифікації для фільтрації Твіттер-стрічки.

```
import tweepy
from tweepy import OAuthHandler, Stream
from tweepy.streaming import StreamListener
from threading import Thread
import json
from textblob import TextBlob
import pandas as pd
from sklearn.externals import joblib
import tornado.ioloop
import tornado.web
import tornado.websocket
from tornado.options import define, options

used_features = ['screen_name_binary', 'name_binary', 'description_binary', 'status_binary',
'listed_count_binary',
        'verified', 'followers_count', 'friends_count', 'statuses_count']
bag_of_words_bot = 'bot', 'b0t'

define("port", default=8888, type=int)

websockets = []

class IndexHandler(tornado.web.RequestHandler):
  def data_received(self, chunk):
    pass
  def get(self):
    self.render("index.html")


class WebSocketHandler(tornado.websocket.WebSocketHandler):
  def data_received(self, chunk):
    pass
  def open(self, *args):
    print("New connection")
    websockets.append(self)
    self.write_message("Welcome!")
  def on_message(self, message):
    print("New message {}".format(message))
```

```
        self.write_message(message.upper())

    def on_close(self):
        websockets.remove(self)
        print("Connection closed")

app = tornado.web.Application([
    (r'/', IndexHandler),
    (r'/ws/', WebSocketHandler),
])

def contains(str):
    for word in bag_of_words_bot:
        if word in str.lower():
            return True
    return False

class MyListener(StreamListener):
    def on_data(self, data):
        try:
            twitt = json.loads(data)
            print("Message " + twitt["text"])
            user = twitt["user"]
            print("Username " + user["name"])
            data = [contains(user["screen_name"]),
                    contains(user["name"]),
                    contains(user["description"]),
                    (user["statuses_count"] > 0),
                    (user["listed_count"] < 20000),
                    user["verified"],
                    user["followers_count"],
                    user["friends_count"],
                    user["statuses_count"]]
            input_data = pd.DataFrame([data])
            result = final_model.predict(input_data)
            print("Bot: " + str(result[0]))
            print("Sentiment: " + str(get_sentiment(twitt["text"])))
            print()
            for socket in websockets:
                socket.write_message(twitt)
            return True
        except Exception as e:
            print(e)
```

```
    def on_error(self, status):
        print(status)
        return True

def listen(tags=['#android', "#java", "#kotlin", "#spring", ""]):
    twitter_stream = Stream(auth, MyListener())
    twitter_stream.filter(track=tags)

def get_sentiment(text):
    textblob = TextBlob(text)
    if textblob.detect_language() != 'en':
        textblob = textblob.translate(to="en")
    return textblob.sentiment

if __name__ == '__main__':
    consumer_key =os.environ['CONSUMER_KEY']
    consumer_secret = os.environ['CONSUMER_SECRET']
    access_token = os.environ['ACCESS_TOKEN']
    access_secret = os.environ['ACCESS_SECRET']

    auth = OAuthHandler(consumer_key, consumer_secret)
    auth.set_access_token(access_token, access_secret)

    api = tweepy.API(auth)

    final_model = joblib.load("model.sav")

    thread = Thread(target=listen)
    thread.start()

    app.listen(options.port)
    tornado.ioloop.IOLoop.instance().start()
```